

## Self-gravitating accretion discs

GIUSEPPE LODATO<sup>(1)</sup>

<sup>(1)</sup> *Department of Physics and Astronomy, University of Leicester, University Road, LE1 7RH, Leicester, UK*

(ricevuto ?)

### Summary. —

I review recent progresses in the dynamics and the evolution of self-gravitating accretion discs. Accretion discs are a fundamental component of several astrophysical systems on very diverse scales, and can be found around supermassive black holes in Active Galactic Nuclei (AGN), and also in our Galaxy around stellar mass compact objects and around young stars. Notwithstanding the specific differences arising from such diversity in physical extent, all these systems share a common feature where a central object is fed from the accretion disc, due to the effect of turbulence and disc instabilities, which are able to remove the angular momentum from the gas and allow its accretion. In recent years, it has become increasingly apparent that the gravitational field produced by the disc itself (the disc's self-gravity) is an important ingredient in the models, especially in the context of protostellar discs and of AGN discs. Indeed, it appears that in many cases (and especially in the colder outer parts of the disc) the development of gravitational instabilities can be one of the main agents in the redistribution of angular momentum. In some cases, the instability can be strong enough to lead to the formation of gravitationally bound clumps within the disc, and thus to determine the disc fragmentation. As a result, progress in our understanding of the dynamics of self-gravitating discs is essential to understand the processes that lead to the feeding of both young stars and of supermassive black holes in AGN. At the same time, understanding the fragmentation conditions is important to determine under which conditions AGN discs would fragment and form stars and whether protostellar discs might form giant gaseous planets through disc fragmentation.

PACS 95.30.Lz – Hydrodynamics.

PACS 97.10.Gz – Accretion and accretion disks.

## 1. – INTRODUCTION

Disc-like or flattened geometries are very common in astrophysics, from the large scale of spiral galaxies down to the small scales of Saturn's rings. In both these examples, the discs are characterized by a prominent structure either in the form of a spiral structure (in the case of galaxy discs) or in the form of gaps and spirals on small scales (in the

case of Saturn's rings). The dynamics underlying the development of such structures is determined by the propagation of density waves, and the important role of the disc's self-gravity in their development has been clearly recognized (see, for example, [1]). In the above examples the system is either *collisionless*, such as in the case of spiral galaxies, where the dominant component is constituted of stars, or *particulate*, such as in the case of Saturn's rings, where the dominant components (mainly rocks and pebbles) do undergo inelastic collisions, but the system cannot be simply described in terms of hydrodynamics. In the last thirty years increasing attention has been given to *fluid* discs, where the dynamically active component is gaseous. Here, dissipative effects, associated with friction or viscosity, can significantly affect the dynamics of the disc. Through dissipation the fluid elements in the disc can lose their energy or, more fundamentally, their angular momentum, as I describe in more detail below, and can fall towards the bottom of the potential well, hence accreting on to a central gravitating body. Such a system, where a disc feeds a central object through accretion, under the effect of viscous forces, is called an **accretion disc**.

Accretion discs are found in very different contexts and over a wide range of physical scales. On the largest scale, they are one of the main physical ingredient to power the central engine of Active Galactic Nuclei (AGN), through accretion on to a supermassive black hole. The mass of the central black hole ranges from  $10^6 M_\odot$  to  $10^9 M_\odot$  and the accretion discs can extend out to large distances, of the order of about 1 pc, where rotating gaseous discs have often been detected through water maser emission [2-4]. On the galactic scale, accretion discs around compact objects, such as white dwarfs, neutron stars and stellar mass black holes, are often found in binary systems, where the compact object (with a mass of the order of  $1 M_\odot$ ) is fed by material outflowing from the companion. Historically, this has been one of the first contexts where accretion disc theory has found widespread application (a detailed review with an emphasis on accretion in galactic binary system and AGN can be found in the textbook by Frank, King and Raine [5]). Another class of objects where accretion discs play an important role are Young Stellar Objects (YSO). Here, the central accreting object is a young protostar, which receives most of its mass from a surrounding protostellar discs. In this case, protostellar discs play also another key role, as the site where planet formation takes place. Understanding the dynamics of the disc and the development of the various instabilities that might occur in it, and possibly lead to turbulence, is therefore essential if one wants to understand some properties of our own Solar system, as well as those of extra-solar planets. A thorough description of the current observational state of protostellar discs can be found, for example, in the recent Protostars and Planets V book [6].

As for the case of spiral galaxies, also for accretion discs the effects of the disc self-gravity can be very important. Indeed, also in this case the development of gravitational instabilities can lead to the formation of grand design spiral structures, which deeply affect the structure and the dynamics of the disc. The analysis of such instabilities naturally reveals several similarities between the case of spiral galaxies and that of accretion discs, but also some important differences. First of all, as mentioned above, accretion discs are fluid and, as such, they are intrinsically dissipative. Indeed, the evolution of gravitationally unstable accretion discs, as I will discuss below, is strongly dependent on the gas cooling rate. On the other hand, gravitationally unstable discs of stars, in the absence of any gas, have a natural tendency to dynamically heat up. Thermal energy, or disordered kinetic energy, plays a key role in stabilizing self-gravitating discs. However, while for stellar discs this energy is mostly in the form of disordered stellar motions, with the possibility of an anisotropic velocity dispersion, for gaseous discs the disordered

kinetic energy content is mostly in the form of thermal energy. Furthermore, for gaseous discs it is expected that the role of resonances is much reduced with respect to a collisionless system. Finally, while for galaxy discs the disc mass can be a substantial fraction of the total mass of the system and therefore give a large contribution to the gravitational potential, in the case of accretion discs the disc mass is often much smaller than the mass of the central object (with some important exceptions, described below).

Historically, accretion disc theory has concentrated on the non self-gravitating case, the effects of self-gravity having only been discussed occasionally [7-10]. In the last ten years, on the other hand, the important role of the disc self-gravity has been clearly recognized, partly due to improved observations, which have shown that in several observed systems (on all scales, from AGN to protostars) the disc mass can be high enough to have a dynamical role, and partly due to the increased computational resources, which have allowed a detailed numerical investigation of the development of gravitational instabilities in the non-linear regime. However, the early papers mentioned above already reveal the main directions around which research in this field would have later developed. Currently, most of the interest revolves around the three issues of self-regulation [7], disc fragmentation [8] and angular momentum transport induced by gravitational instabilities [9,10]. This again reveals a difference with respect to the galaxy discs case, where most of the attention has been traditionally given to the morphology induced by self-gravity.

In this paper, I present an overview of the dynamics of self-gravitating accretion discs. Accretion disc theory is a very vast topic and has been treated extensively elsewhere [5,11], so I will not repeat it here. I will rather summarize the most salient general features, with particular emphasis on those features which are most affected by the disc self-gravity. Similarly, a detailed description of all the different astrophysical systems where self-gravitating accretion discs play a role (protostellar discs, AGN,...) is obviously beyond the scope of this review. The structure of the paper is as follows. In section 2 I describe the basic equations that determine the evolution of accretion discs and the main effects of the disc self-gravity. In section 3 I discuss the development of gravitational instabilities, and present some models which incorporate in a simple way the salient physics behind their development. In section 4 I review the current state of numerical simulations of gravitationally unstable gaseous discs. In section 5 I focus on the transport properties induced by gravitational instabilities, while in section 6 I describe the related issue of disc fragmentation. Finally, in the last two sections I will present some examples of observed systems where the concepts described in this paper find a natural application, and in particular, in section 7 I will discuss the impact of self-gravity in the dynamics of protostellar discs with some application related to planet formation. In section 8 I focus on the impact of self-gravity in AGN discs also showing how self-gravitating accretion discs might have played an important role at high red-shifts, by allowing the formation of the seeds of supermassive black holes.

## 2. – SELF-GRAVITATING ACCRETION DISC DYNAMICS

**2.1. The thin disc approximation.** – One of the main properties of accretion discs is that they are often thin. This means that the typical lengthscale in the vertical direction, the disc thickness  $H$ , is much smaller than the radial distance from the central object  $R$ . For example, AGN discs are generally very thin, with aspect ratios  $H/R \approx 0.001 - 0.01$ , while protostellar discs are comparably a bit thicker, with  $H/R \approx 0.1$ . As we will discuss below, this has a significant impact on the conditions under which the disc is self-gravitating, and hence this intrinsic difference between the AGN case and the protostellar

case will be reflected in a significantly different behaviour of these two kind of systems in the self-gravitating regime.

The thin disc condition allows us to treat the disc, to a first approximation, as infinitesimally thin, and to introduce a *small quantity*, the aspect ratio  $H/R \ll 1$ . This implies that most of the equations we are going to use can be integrated in the vertical direction, and rather than dealing with quantities *per unit volume* (such as the density  $\rho$ ), we will deal instead with quantities *per unit surface* (such as the surface density  $\Sigma = \int_{-\infty}^{\infty} \rho dz$ ). When “volume” quantities are needed (for example the viscosity  $\nu$ ), these will generally be understood as vertically averaged.

The fundamental ordering of lengthscales  $H/R \ll 1$  is also related to a similar ordering in terms of velocities. Indeed, one can show (see section 2.4) that this ordering is equivalent to requiring that the sound speed  $c_s$  is much smaller than the rotational velocity  $v_\phi$ . It is useful to anticipate another important relation between the relevant velocities, derived by requiring that accretion takes place on a long timescale. This condition implies that the radial velocity  $v_R$  should be smaller than both the sound speed and the rotational speed. We can therefore summarize these relations as:

$$(1) \quad v_R \ll c_s \ll v_\phi.$$

Finally, it is worth adding a word of caution on the applicability of the thin disc condition. Indeed, in some cases the disc might not be very thin. This happens for example, in the hotter inner regions of AGN discs. Protostellar discs, as mentioned above, are generally relatively thick. One should therefore be cautious when using the thin disc approximation, especially when considering quantities averaged in the vertical direction.

**2.2. Basic equations.** – The evolution of accretion discs is determined by the basic equations of viscous fluid dynamics: the continuity equation and Navier-Stokes equations. Consider an accretion disc rotating around a central object of mass  $M$ . Given the geometry of the problem, we adopt cylindrical polar coordinates centred on the central object and we assume that to first order the disc is axisymmetric, so that no quantity depends on the azimuthal angle  $\phi$ . Consider then a disc with surface density  $\Sigma(R, t)$ . In cylindrical coordinates, the continuity equation (integrated in the vertical direction, following the thin disc approximation) reads

$$(2) \quad \frac{\partial \Sigma}{\partial t} + \frac{1}{R} \frac{\partial}{\partial R} (R \Sigma v_R) = 0.$$

To determine the velocity  $\mathbf{v}$ , we use the momentum equation. For an inviscid fluid this takes the form of Euler’s equation. For the moment, we consider the full three-dimensional (i.e. not vertically integrated) equation

$$(3) \quad \rho \left[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -\nabla P - \rho \nabla \Phi.$$

On the right-hand side there are the various forces acting on the fluid. First of all, we have the term describing pressure forces, where  $P$  is the pressure and  $\rho$  is the density. We then have gravity, represented by the last term on the right-hand side, where  $\Phi$  is the gravitational potential. This in general includes both the contribution of the central

point mass and that of the disc. A detailed description of these different contributions is provided in the next section. For an infinitesimally thin disc, we can obtain the potential associated with the gas surface density  $\Sigma$  (the ‘self-gravity’) from the solution to Poisson’s equation:

$$(4) \quad \nabla^2 \Phi_{\text{sg}} = 4\pi G \Sigma \delta(z),$$

where  $\delta(z)$  is the Dirac  $\delta$ -function, indicating that the disc density is confined to the midplane.

For an accretion disc, however, Euler equation is not sufficient, because it does not include the important terms due to viscosity. Including viscous forces, the relevant momentum equation is the Navier-Stokes equation. This reads

$$(5) \quad \rho \left[ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -(\nabla P - \nabla \cdot \sigma) - \rho \nabla \Phi.$$

The second term on the right hand side in equation (5) contains the stress tensor  $\sigma$  and describes the effect of viscous forces. This term plays a very important role in accretion disc dynamics. The nature of the disc viscosity and of the stress tensor  $\sigma$  has been traditionally one of the main unsolved theoretical issues in accretion disc dynamics. In its simplest form, it can be assumed to be given by classical shear viscosity, so that the only non-vanishing component of  $\sigma$  in a circular shearing flow is the  $R\phi$  component, proportional to the rate of strain  $R\Omega'$  (where  $\Omega' = d\Omega/dR$ ):

$$(6) \quad \sigma_{R\phi} = \eta R \frac{d\Omega}{dR},$$

where  $\Omega = v_\phi/R$  is the angular velocity and we have introduced the shear viscosity coefficient  $\eta$ . Equation (5) has three components in the radial, vertical and azimuthal directions, respectively, and each one defines some important properties of accretion discs. In the following three sections we will examine in turn each of these three equations.

**2.3. Radial equilibrium: centrifugal balance.** – Let us first consider the radial component of equation (5). Here, the ordering of velocities described above turns out to be particularly useful. In particular, on the left-hand side, the first term, that contains the radial velocity  $v_R$  is negligible with respect to the second term, which (when using the appropriate expression for the differential operators in cylindrical coordinates) gives rise to the centrifugal term  $v_\phi^2/R$ . On the right-hand side, the viscous term vanishes and the pressure term can be obtained using the equation of state. If the gas is barotropic (i.e. if pressure only depends on density), the sound speed is simply defined as:

$$(7) \quad c_s^2 = \frac{dP}{d\rho}.$$

We then have:

$$(8) \quad \frac{1}{\rho} \frac{\partial P}{\partial R} = \frac{c_s^2}{\rho} \frac{\partial \rho}{\partial R} \sim \frac{c_s^2}{R}.$$

Since  $c_s \ll v_\phi$ , also this term is second order with respect to the leading term. The only term on the right hand side able to balance the centrifugal force is therefore gravity. We then have, to first order:

$$(9) \quad \frac{v_\phi^2}{R} \simeq \frac{\partial \Phi}{\partial R}.$$

For accretion discs the gravitational field is generally dominated by the central compact object, but in some cases the disc self-gravity can also produce a sizable effect. The central object contribution to the field is simply given by

$$(10) \quad \frac{\partial \Phi}{\partial R} = \frac{GM}{R^2}$$

Thus, in the limit where we neglect self-gravity, the disc rotation is Keplerian, with  $v_\phi = v_K \equiv \sqrt{GM/R}$  and where the angular velocity is given by  $\Omega = \Omega_K \equiv \sqrt{GM/R^3}$ .

The disc contribution to the gravitational field depends on the matter distribution through Poisson's equation. For an infinitesimally thin disc, the relation between  $\Sigma$  and the disc gravitational field can be written, for example, using the complete elliptic integrals of the first kind,  $K$  and  $E$

$$(11) \quad \frac{\partial \Phi_{\text{disc}}}{\partial R}(R, z) = \frac{G}{R} \int_0^\infty \left[ K(k) - \frac{1}{4} \left( \frac{k^2}{1-k^2} \right) \left( \frac{R'}{R} - \frac{R}{R'} + \frac{z}{RR'} \right) E(k) \right] \sqrt{\frac{R'}{R}} k \Sigma(R') dR',$$

where  $k^2 = 4RR'/[(R+R')^2 + z^2]$  [12, 13]. The above equation in general describes a complicated relation between  $\Sigma$  and the gravitational field, which can be obtained numerically. However, there are a few cases where the relation is particularly simple and can be derived analytically. One interesting case is the one described by Mestel [14], where  $\Sigma = \Sigma_0 R_0/R$  and the disc extends out to infinity. In this case the gravitational field at the disc midplane takes a particularly simple form:

$$(12) \quad \frac{\partial \Phi_{\text{disc}}}{\partial R}(R) = 2\pi G \Sigma(R).$$

In the extreme limit, where the central object contribution is negligible and a Mestel disc dominates the gravitational field [15], the rotation curve is thus flat:

$$(13) \quad v_\phi^2 = R \frac{\partial \Phi_{\text{disc}}}{\partial R}(R) = 2\pi G \Sigma(R) R = 2\pi G \Sigma_0 R_0,$$

and the angular velocity  $\Omega \propto 1/R$ . Interestingly, as described in more detail in section 3.3 below, it turns out that a Mestel disc is the natural solution to the accretion disc equations in a steady state, in the limit of strongly self-gravitating discs [15]. Now, while in actual accretion discs the central object mass is not expected to be negligible, it is interesting to note that in many cases observations of protostellar discs imply surface densities  $\Sigma \propto 1/R$  [16]).

Having obtained the rotation curve in the two limiting cases of negligible disc self-gravity and negligible central object gravity, we can now ask under which conditions do we expect the different contributions to dominate. Comparing Eq. (10) to Eq. (12),

we see that in order for the disc contribution to be comparable to the central object we require

$$(14) \quad M_{\text{disc}}(R) \approx 2\pi\Sigma(R)R^2 \approx M,$$

that is, we require disc masses of the order of the central object mass. This is confirmed by detailed models of self-regulated discs [13]. In most cases, the accretion disc mass is much lower than the central object and so the Keplerian approximation is generally appropriate. However, there are a few important exceptions where non-Keplerian rotation has been observed and can be interpreted in terms of relatively massive discs. This occurs for example for some protostellar discs around massive stars [17, 18] and also in some AGN discs, such as in the nucleus of the active galaxy NGC 1068 [3, 19]. We defer a more detailed description of such cases to section 7 and 8 below.

The above derivation is only valid to first approximation, since we have neglected the pressure forces in the radial direction. This is generally a good approximation, but in some cases the corrections due to pressure effects might play an important role. In particular, this is the case when one considers the dynamics of small solid bodies within the disc (a very important component involved in the process of planet formation, as discussed below in section 7.5). The solids are not subject to pressure and would thus move on exactly Keplerian orbits, therefore generating a small velocity difference with respect to the gas (in a process somewhat analogous to the so called asymmetric drift in stellar dynamics), which in turn determines a very fast migration of the solids [20–22]. Let us then calculate this correction. The radial equation of motion, including pressure terms is:

$$(15) \quad \frac{v_\phi^2}{R} = \frac{1}{\rho} \frac{\partial P}{\partial R} + \frac{v_{\text{circ}}^2}{R},$$

where we have indicated with  $v_{\text{circ}}^2 = R(d\Phi/dR)$  the circular velocity in the absence of pressure. If we introduce the sound speed  $c_s$  and make the simple assumption that the density  $\rho$  has a power law dependence on radius, with power law index  $-\beta$  (so that  $\rho \propto R^{-\beta}$ ), we obtain:

$$(16) \quad v_\phi = v_{\text{circ}} \left[ 1 - \beta \left( \frac{c_s}{v_{\text{circ}}} \right)^2 \right]^{1/2}.$$

In cases where the disc is hot, in the sense that the sound speed is non negligible with respect to  $v_{\text{circ}}$ , then pressure effects will give a sizable contribution to the rotation curve.

**2.4. Vertical structure: hydrostatic equilibrium.** – We now consider the vertical component of equation (5). Here, since the velocity in the vertical direction is very small (the disc being confined to the equatorial plane), we can neglect the left-hand side of the equation altogether. The viscous force vanishes as well, since the only non-zero component of the stress is in the  $R\phi$  direction. We are therefore left with just two terms to balance: gravitational force and pressure force in the vertical direction. Such a situation, where gravity is balanced by pressure, is called “hydrostatic balance”. The equation reads:

$$(17) \quad \frac{1}{\rho} \frac{\partial P}{\partial z} = -\frac{\partial \Phi}{\partial z}.$$

We again first consider the case of non self-gravitating discs. In this case, the vertical component of gravity is simply:

$$(18) \quad -\frac{GM}{r^2} \frac{z}{r} \sim -\frac{GM}{R^2} \frac{z}{R},$$

where  $r$  is the spherical radius and the approximation is valid for small  $z$ . We can then rewrite Equation (17) as:

$$(19) \quad \frac{c_s^2}{\rho} \frac{\partial \rho}{\partial z} = -\frac{GMz}{R^3}.$$

The solution to this equation is straightforward if the sound speed is independent of  $z$ . The vertical density profile in this case turns out to be a Gaussian:

$$(20) \quad \rho(z) = \rho_0 \exp \left[ -\frac{GMz^2}{2R^3c_s^2} \right] = \rho_0 \exp \left[ -\frac{z^2}{2H_{\text{nsg}}^2} \right],$$

where  $\rho_0$  is the midplane density and we have introduced a typical vertical scale-length in the non-self-gravitating case  $H_{\text{nsg}}$ :

$$(21) \quad H_{\text{nsg}} = \frac{c_s}{\Omega_K}.$$

It is also worth introducing a slightly modified scale-length<sup>(1)</sup>  $H_* = \sqrt{\pi/2} H_{\text{nsg}}$ , so that the surface density  $\Sigma = \int_{-\infty}^{\infty} \rho dz$  is related to the midplane density by  $\Sigma = 2\rho_0 H_*$ . Note the simple relation between the thickness, the sound speed and the angular velocity. The disc aspect ratio is then:

$$(22) \quad \frac{H_{\text{nsg}}}{R} = \frac{c_s}{v_K},$$

which then demonstrates, as anticipated above, that the requirement that the disc be thin is equivalent to requiring that the disc rotation is highly supersonic, i.e. that  $v_K \gg c_s$ .

Let us now consider the case of a disc dominated by self-gravity. In this case, we simply have to replace the gravitational force on the right hand side of equation (19) with the force produced by a slab of gas with surface density  $\Sigma$ . If the slab is radially homogeneous we have

$$(23) \quad \frac{c_s^2}{\rho} \frac{\partial \rho}{\partial z} = -2\pi G \Sigma(z),$$

where  $\Sigma(z) = \int_{-z}^z \rho(z') dz'$ . The solution of the hydrostatic balance in this case is more difficult but can be done analytically in the case of constant sound speed (such case is

---

<sup>(1)</sup> The small difference between  $H_{\text{nsg}}$  and  $H_*$  is generally neglected in many papers, and the two expressions are both commonly used to indicate the disc thickness in the non-self-gravitating case.



generally referred to as the “self-gravitating isothermal slab”) [23]. The density profile in this case is not Gaussian, but is given by:

$$(24) \quad \rho(z) = \rho_0 \frac{1}{\cosh^2(z/H_{\text{sg}})},$$

where the thickness in the self-gravitating case is:

$$(25) \quad H_{\text{sg}} = \frac{c_s^2}{\pi G \Sigma},$$

where in this case the surface density and the midplane density are related by  $\Sigma = 2\rho_0 H_{\text{sg}}$ .

Now, we can ask the same question that we asked above for the radial contribution to the gravitational field. How massive has to be the disc, in order for self-gravity to significantly affect its vertical structure? Note that the vertical component of the gravitational field produced by the disc is of the order of  $2\pi G \Sigma \sim GM_{\text{disc}}/R^2$ . On the other hand, the vertical component of the star’s gravitational field is smaller than the radial ( $GM/R^2$ ) by a factor  $H_{\text{nsg}}/R$  (cf. Eq. (18)) and thus [7] the vertical structure of the disc is affected by self-gravity already when the disc mass is of the order of:

$$(26) \quad \frac{M_{\text{disc}}}{M} \approx \frac{H_{\text{nsg}}}{R} \ll 1.$$

Thus, for thin discs the vertical structure of the disc is affected by the disc self-gravity, even when the disc mass is much smaller than the central object mass. Another way of looking at the same question is to compare the two expressions for the thickness obtained above:

$$(27) \quad \frac{H_{\text{sg}}}{H_{\text{nsg}}} = \frac{c_s \Omega_K}{\pi G \Sigma}.$$

The two expressions become comparable when the dimensionless parameter on the right hand side is of order unity. It is then easy to see that this occurs when  $M_{\text{disc}}/M \approx H_{\text{nsg}}/R$ . The parameter on the right (to within factors of order unity) is nothing else than the well known  $Q$  parameter that controls the development of gravitational instabilities in the disc [24, 25]. These will be discussed at length starting from section 3 below.

It is also possible to solve the vertical hydrostatic balance equation in the combined case where both the central object and the disc contribute to the gravitational field. In this case, we have:

$$(28) \quad \frac{1}{\rho} \frac{\partial \rho}{\partial z} = - \frac{2\pi G \Sigma(z) + \Omega_K^2 z}{c_s^2}.$$

This equation does not have an analytical solution and has to be solved numerically. However, an approximate but accurate solution can be obtained also in this case [13], with the thickness given by

$$(29) \quad H = \frac{H_{\text{sg}}}{2} \left( \frac{H_\star}{H_{\text{sg}}} \right)^2 \left[ \sqrt{1 + 4 \left( \frac{H_{\text{sg}}}{H_\star} \right)^2} - 1 \right],$$

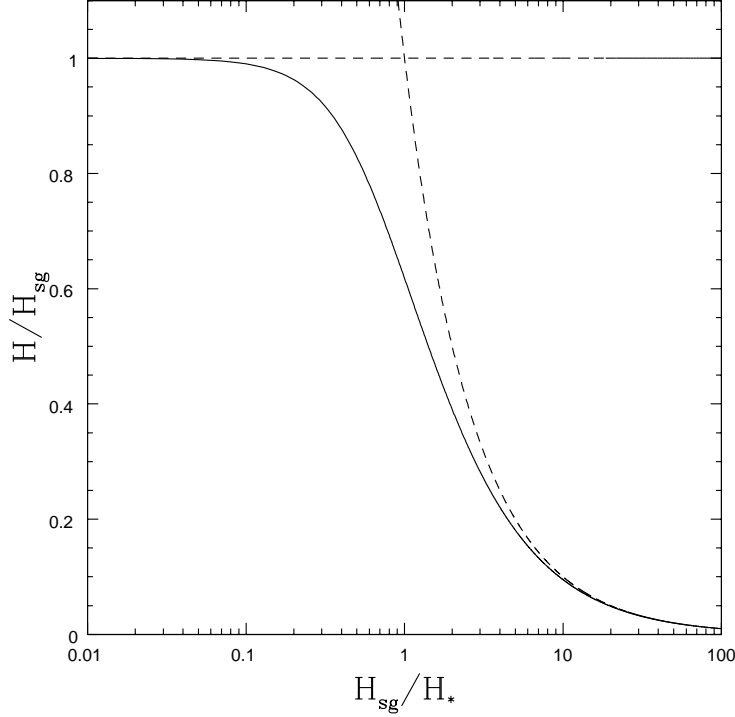


Fig. 1. – Disc thickness from on Eq. (29) as a function of  $H_{\text{sg}}/H_*$ . The two dashed lines indicate the limiting cases  $H = H_{\text{sg}}$  and  $H = H_*$ .

which in this form holds for Keplerian rotation. Further correction terms for non-Keplerian cases can be easily incorporated in the above expression [13]. Note that Eq. (29) correctly reproduces the limiting cases of fully self-gravitating discs,  $H_{\text{sg}} \ll H_*$ , where  $H \approx H_{\text{sg}}$  and that of non-self-gravitating disc,  $H_* \ll H_{\text{sg}}$ , where  $H \approx H_*$ . In the following we will simply use  $H$  to indicate the disc thickness. Figure 1 shows the thickness calculated from Eq. (29) as a function of  $H_{\text{sg}}/H_*$ , along with the two limiting cases  $H = H_{\text{sg}}$  and  $H = H_*$ . Note that when  $H_{\text{sg}} \approx H_*$  (which, as I discuss below, is a common case for self-gravitating discs), we have  $H \approx 0.62H_{\text{sg}} = 0.62H_*$ .

**2.5. Angular momentum conservation: the issue of ‘anomalous’ viscosity.** – Let us finally discuss the last component of Navier-Stokes equation, in which the viscous term plays an important role. First, we integrate Eq. (5) in the vertical direction:

$$(30) \quad \Sigma \left( \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right) = -(\nabla P - \nabla \cdot T) - \rho \nabla \Phi,$$

where  $T$  is the vertical integral of the stress tensor (Eq. (6)), whose only non-vanishing component is

$$(31) \quad T_{R\phi} = \nu \Sigma R \frac{d\Omega}{dR}.$$

In Eq. (31) the vertically averaged kinematic viscosity  $\nu$  is defined through

$$(32) \quad \nu\Sigma = \int_{-\infty}^{\infty} \eta dz.$$

Combining the  $\phi$  component of Eq. (30) with continuity equation (Eq. (2)) and after a little algebra, we obtain the expression for angular momentum conservation for a viscous disc:

$$(33) \quad \frac{\partial}{\partial t} (\Sigma R v_\phi) + \frac{1}{R} \frac{\partial}{\partial R} (R v_R \Sigma R v_\phi) = \frac{1}{R} \frac{\partial}{\partial R} (\nu \Sigma R^3 \Omega'),$$

where  $\Omega' = d\Omega/dR$ . The physical interpretation of the above expression is readily apparent. The left hand side is the Lagrangian derivative of the angular momentum per unit mass  $\Sigma R v_\phi$ , while the right hand side is the torque exerted by viscous forces.

With the help of the continuity equation (Eq. (2)), we can obtain the radial velocity from equation (33):

$$(34) \quad v_R = \frac{1}{R\Sigma(R^2\Omega)'} \frac{\partial}{\partial R} (\nu \Sigma R^3 \Omega'),$$

which can be inserted back in equation (2) to finally give:

$$(35) \quad \frac{\partial \Sigma}{\partial t} = -\frac{1}{R} \frac{\partial}{\partial R} \left[ \frac{1}{(R^2\Omega)'} \frac{\partial}{\partial R} (\nu \Sigma R^3 \Omega') \right],$$

which is the master equation that describes the evolution of an accretion disc. In general, in Eq. (35),  $\nu$  can depend on both radius and  $\Sigma$ , and even  $\Omega$ , for self-gravitating discs, may be a function of  $\Sigma$ , thus making Eq. (35) a non-linear evolutionary equation. However, if  $\Omega$  and  $\nu$  are independent of  $\Sigma$ , then Eq. (35) is a diffusion equation and a simple analysis of its structure shows that the disc surface density evolves on a timescale of the order of

$$(36) \quad t_{\text{visc}} = \frac{R^2}{\nu}.$$

In particular, in the important case of Keplerian rotation, the diffusion equation has the following form:

$$(37) \quad \frac{\partial \Sigma}{\partial t} = \frac{3}{R} \frac{\partial}{\partial R} \left[ R^{1/2} \frac{\partial}{\partial R} (\nu \Sigma R^{1/2}) \right].$$

The above discussion shows the extremely important role of viscosity in the evolution of accretion discs. However, unfortunately, the ultimate physical origin of viscosity is also the major unsolved problem in accretion disc theory. As it was soon realized, standard kinetic viscosity due to collisions between gas molecules is far too low to account for the transport of angular momentum needed in observed accretion discs. In order to see this, let us consider the viscous timescale introduced above,  $t_\nu = R^2/\nu$ . It is convenient to express this quantity in units of the dynamical timescale  $t_{\text{dyn}} = \Omega^{-1}$ :

$$(38) \quad \frac{t_\nu}{t_{\text{dyn}}} = \frac{R^2 \Omega}{\nu}.$$

Note that the ratio above is nothing else than the Reynolds number  $Re$  of the flow. Standard, collisional viscosity can be expressed as the product between the typical random velocity of molecules (that will be of order of the sound speed  $c_s$ ) times the collisional mean free path  $\lambda = 1/(n\sigma_{\text{coll}})$ , where  $n$  is the number density of the gas and  $\sigma_{\text{coll}}$  is the collisional cross-section. We then have:

$$(39) \quad \lambda = \frac{1}{n\sigma_{\text{coll}}} = \frac{\mu m_p}{\rho \sigma_{\text{coll}}} = \left( \frac{2\mu m_p}{\Sigma \sigma_{\text{coll}}} \right) H,$$

where  $m_p \approx 10^{-24}\text{g}$  is the proton mass,  $\mu$  is the average molecular weight (we can take it to be  $\approx 2$ , for simplicity) and where  $\rho$  and  $\Sigma$  are the usual volume and surface density of the disc, respectively. Now, substituting  $\nu = \lambda c_s$  in equation (38), we get:

$$(40) \quad \frac{t_\nu}{t_{\text{dyn}}} = \left( \frac{\Sigma \sigma_{\text{coll}}}{2\mu m_p} \right) \left( \frac{H}{R} \right)^{-2}.$$

To give some ideas of the numbers involved, let us assume that the collisional cross section is simply of the order of the size of an hydrogen molecule  $\sigma_{\text{coll}} \approx 10^{-16} \text{ cm}^2$ . In order to give a rough estimate of the disc surface density let us consider a typical protostellar disc, and assume (to be conservative) that the disc mass is  $\approx 0.005 M_\odot$  and that the disc size is  $\approx 50 \text{ AU}$ , which gives us  $\Sigma \approx 0.005 M_\odot / (50 \text{ AU})^2 \approx 10 \text{ g/cm}^2$ . Inserting this numbers in Eq. (40), and also assuming that  $H/R \approx 0.1$ , we get  $t_\nu/t_{\text{dyn}} \approx 10^{11}$ . Since the dynamical timescale for protostellar discs is of the order of a few years, the above estimates would lead to the conclusion that in order to accrete a disc of mass  $0.005 M_\odot$  from a distance of  $50 \text{ AU}$ , it would take much longer than the Hubble time! Clearly, the magnitude of viscosity must be much larger than the simple collisional one estimated above. Similar estimates can also be done for discs in other contexts, such as AGN and galactic binaries, where the gas is expected to be ionized and the transport coefficient will be mostly determined by plasma processes [5]

However, the very small magnitude of collisional viscosity offers a possible way out from the apparent paradox. Indeed, as remarked above, the ratio of the viscous timescale to the dynamical one is also equal to the Reynolds number of the flow. The fact that collisional viscosity gives such a large estimate of this ratio also implies that the Reynolds number of the accretion flow is correspondingly large. Now, it is well known that for high Reynolds numbers a flow is subject to the development of turbulence and we should therefore expect that the flow in an accretion disc should be highly turbulent. In these conditions, viscosity can be much higher because in this case angular momentum is exchanged not through collisions of individual gas molecules, but by the mixing of fluid elements moving around in the disc due to turbulence. The typical length-scale of such motion can be several orders of magnitude larger than the collisional mean free path and consequently transport becomes much more effective. The question is then how to characterize such an ‘anomalous’ viscosity. For decades, modelers have relied on a very simple and successful parameterization of viscosity in terms of an unknown dimensionless parameter, called  $\alpha$  [26]. The argument is very simple and draws from the fact that the stress tensor has the physical dimension of a pressure, that is a density times the square of a velocity. The simplest assumption is then to take the stress tensor to be just proportional to the (vertically integrated) pressure  $\Sigma c_s^2$ :

$$(41) \quad T_{R\phi} = \frac{d \ln \Omega}{d \ln R} \alpha \Sigma c_s^2,$$

where  $d \ln \Omega / d \ln R$  ( $\sim -3/2$  for a quasi Keplerian rotation curve) is just a number (to remind us that the viscous stress is proportional to the rate of strain), and where  $\alpha$  is the proportionality factor between the stress tensor and the pressure.

Another way of expressing the  $\alpha$ -prescription is by considering the kinematical viscosity  $\nu$ . Indeed, Eq. (41) is equivalent to:

$$(42) \quad \nu = \alpha c_s H.$$

This form of the  $\alpha$ -prescription offers a simple way to put some constraints on  $\alpha$ . The magnitude of turbulent viscosity is given roughly by  $\nu \sim \hat{v} l$ , where  $l$  is the typical size of the largest eddies in the turbulent pattern and where  $\hat{v}$  is the typical turbulent velocity. Now, it is unlikely that the turbulence will be highly supersonic, since otherwise it would be easily dissipated through shocks and we thus have  $\hat{v} \lesssim c_s$ . An upper limit to the size of the largest eddies  $l$  is obviously given by the disc thickness  $H$  (if we consider an isotropic turbulence). These two upper limits, taken together, clearly imply that  $\alpha < 1$ . Typically quoted values for  $\alpha$ , derived from comparing theoretical evolutionary models to observations range from  $\alpha \approx 0.01$  [27] for protostellar discs to  $\alpha \approx 0.1$  for galactic binaries [28], although in some more uncertain cases (such as for the outbursting protostellar systems called FU Orionis objects), the models tend to indicate a somewhat smaller value  $\alpha \approx 10^{-3}$  [29, 30].

Note that the arguments above do not necessarily imply that the stress tensor is strictly proportional to the gas pressure. Rather they merely indicate that we may expect the ratio of the two to be smaller than unity. While in many cases the parameter  $\alpha$  is taken to be a constant, it obviously does not need to be so. In this respect, the  $\alpha$ -prescription is not a theory of viscosity in accretion discs, since it only replaces an unknown parameter,  $\nu$ , with another unknown parameter,  $\alpha$ . By using the  $\alpha$ -prescription we simply measure the stress in units of the local pressure, and we realize that in these units the stress should not be larger than unity, if it is driven by local turbulence.

Modern accretion disc theory is thus based on the assumption that ‘anomalous’ transport of angular momentum is provided by turbulent viscosity. Clearly, in order to maintain the turbulence for long enough, energy needs to be continuously injected at the largest scales. This is usually associated with the development of instabilities in the disc. Hydrodynamic, magneto-hydrodynamic and gravitational instabilities have all been considered in this respect and in recent years several numerical simulations have shown that many of the instabilities considered can indeed lead to sustained turbulence. MHD instabilities are often considered to be the most likely cause of ‘anomalous’ viscosity. A detailed discussion of these processes can be found elsewhere [31]. In the context of this paper, I will put more emphasis on the role of gravitational instabilities, and the transport of angular momentum and energy associated with gravitational instabilities will be discussed in detail below in section 5. Gravitational instabilities are the most likely alternative to MHD instabilities in providing a source of transport in accretion discs, and their role is now well established, especially in the context of protostellar discs [32-34].

Before moving on, it is worth pointing out that the term “turbulence” is sometime misused in astronomy, often indicating a simply fluctuating velocity field, which does not necessarily display the typical turbulent cascade from large to small scales. While in some cases simulations of hydrodynamic instabilities do produce turbulent vortices, often the structure consists of larger scale coherent disturbances. This is particularly true for the case of gravitational instabilities, that develop in the form of spiral structures and for which the term “gravo-turbulent”, sometimes used, is probably inappropriate.

**2.6. Simple Keplerian disc models.** – Let us see what kind of evolution is associated with the equation of motion described above in a few simple cases. Let us first consider the case of Keplerian rotation, described by Eq. (37). A particularly simple but instructive configuration is when the viscosity  $\nu$  is a constant, independent of radius. In this case Eq. (37) can be solved analytically [35]. We take as initial condition for the surface density an infinitesimally thin ring of mass  $m$ , whose shape is a  $\delta$  function centered at some radius  $R_0$ :

$$(43) \quad \Sigma(R, t = 0) = \frac{m}{2\pi R_0} \delta(R - R_0)$$

The solution to equation (37) in this case is:

$$(44) \quad \Sigma(x, \tau) = \frac{m}{\pi R_0^2} \frac{x^{-1/4}}{\tau} e^{-(1+x^2)/\tau} I_{1/4} \left( \frac{2x}{\tau} \right),$$

where  $I_{1/4}$  is a modified Bessel function of the first kind,  $x = R/R_0$  and  $\tau = 12\nu t/R_0^2$ . The evolution of the surface density is shown in the left panel of Fig. 2, where the different lines refer (from top to bottom) to different times  $\tau = 0.01, 0.05, 0.1$  and  $0.15$ . This figure illustrates quite clearly why this example is called “the spreading ring”. We see here that, indeed, under the action of viscous forces the disc, rather than merely accreting, spreads both inwards and outwards. This outward spreading is needed in order to conserve angular momentum. Viscosity simply acts on the fluid to redistribute angular momentum between different annuli transporting it from the inside out, the total angular momentum being conserved. In this way, as the inner parts of the disc lose their angular momentum and accrete, the outer parts take up the excess angular momentum and move outward. A detailed analysis of the analytic solution [5, 11] shows that the transition between inward and outward motion occurs at a radius of order  $R_{\text{tr}} \sim t\nu/R_0 \sim R_0(t/t_\nu)$  and is therefore an increasing function of time. Therefore, at late times, only the outermost parts of the disc move outwards, and most of the mass is eventually accreted. For  $t \rightarrow \infty$ , all the mass in the ring is accreted and the angular momentum is transported to infinitely large radii by a negligibly small amount of mass.

Another important class of solution [35, 36] is found in cases where the viscosity has a simple power-law dependence on radius,  $\nu \propto R^b$ . In this case, it is possible to find a self-similar analytical solution, where the initial density profile is such that accretion proceeds almost steadily out to a typical radius  $R_1$  at which the density is exponentially truncated. An example of the evolution in this case is shown in the right-hand panel of Fig. 2, for the case  $b = 1$ . Also in this case, as for the spreading ring, the disc spreads significantly outwards, even if most of the mass is eventually accreted. The transition between inward and outward moving portion of the disc for the self-similar solutions occurs at a transition radius  $R_{\text{tr}} \sim R_1(t/t_\nu(R_1))^{1/(2-b)}$  and therefore increases with time (for the particular case where  $b = 1$  described above, we have that  $R_{\text{tr}}$  increases linearly with time). As the disc empties out, the accretion rate drops steadily.

**2.7. Steady-state solutions.** – In a steady state, the continuity equation takes the form:

$$(45) \quad \dot{M} = -2\pi R v_R \Sigma,$$

where the constant  $\dot{M}$  is the mass accretion rate and the signs have been chosen in such a way that when  $v_R$  is negative (i.e. directed inwards) the accretion rate is positive.

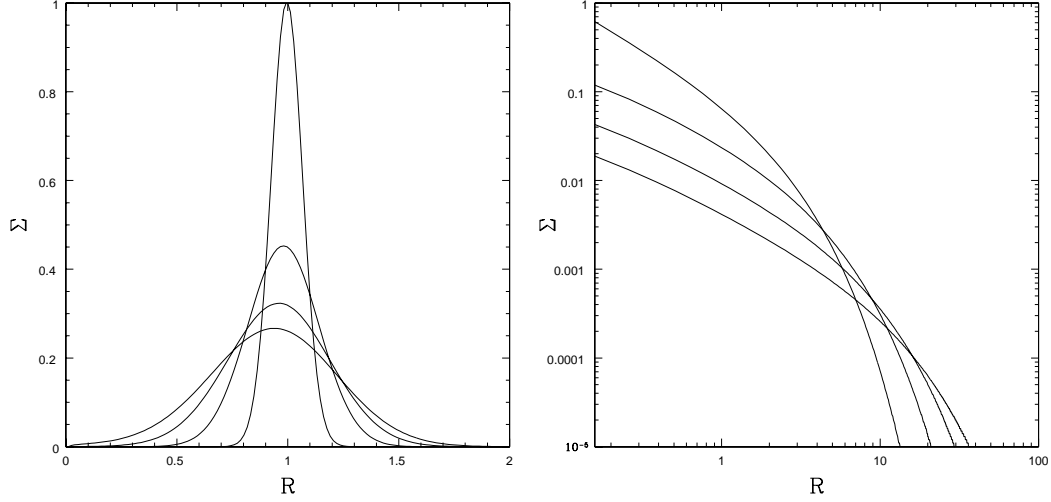


Fig. 2. – Left panel: evolution of the surface density of a Keplerian spreading ring, under the effect of a constant viscosity  $\nu$ . The lines refer (from top to bottom) to  $\tau=0.01, 0.05, 0.1$  and  $0.15$  (see text for the definition of  $\tau$ ). Right panel: evolution of a truncated power-law surface density, in the case where  $\nu \propto R$ . Units are arbitrary.

Analogously, angular momentum conservation becomes:

$$(46) \quad \dot{M}\Omega R^2 - 2\pi\nu\Sigma|\Omega'|R^3 = \dot{J}$$

where  $\dot{J}$  is the constant net flux of angular momentum, and is determined by two contributions: the first term on the left hand side, which indicates the angular momentum advected with the accretion process, and the second term, which indicates the outward flux produced by viscous torques.

$\dot{J}$  is sometimes determined by applying the so-called “no torque” condition, according to which at the disc inner radius  $R_{\text{in}}$  the angular velocity profile flattens (due to the presence of a boundary layer where the disc connects to the central object) so that its gradient and therefore the viscous torque vanish. For Keplerian rotation, this implies  $\dot{J} = \dot{M}(\Omega R^2)_{\text{in}} = \dot{M}\sqrt{GM R_{\text{in}}}$ . Inserting this in equation (46), we obtain:

$$(47) \quad 3\pi\nu\Sigma = \dot{M} \left( 1 - \sqrt{\frac{R_{\text{in}}}{R}} \right).$$

At large radii,  $R \gg R_{\text{in}}$ , where the effects of a finite  $\dot{J}$  are negligible, the surface density and the viscosity satisfy the following simple relation:

$$(48) \quad \dot{M} = \left| \frac{d \ln \Omega}{d \ln R} \right| 2\pi\nu\Sigma,$$

that is, surface density and viscosity are inversely proportional to each other. An interesting consequence of the above relation arises when we use the  $\alpha$ -prescription in the

self-gravitating case, where  $H = c_s^2/\pi G\Sigma$ . Indeed, in this case Eq. (48) becomes:

$$(49) \quad \dot{M} = 2\alpha \frac{c_s^3}{G} \left| \frac{d \ln \Omega}{d \ln R} \right|,$$

which shows that for a self-gravitating disc the accretion rate  $\dot{M}$  is only dependent on  $\alpha$  and on the sound speed  $c_s$ . In particular, a self-gravitating disc with a constant  $\alpha$  in a steady state is approximately isothermal.

**2'8. Temperature profile and spectral energy distribution.** – Let us conclude the analysis of the main properties of accretion discs by considering the energetics associated with the accretion process. One of the most important consequences of accretion is the release of gravitational potential energy as the accreting matter falls into the potential well. The power per unit surface  $D(R)$  dissipated by viscous stresses is given by:

$$(50) \quad D(R) = \nu \Sigma (R\Omega')^2 = \left| \frac{d \ln \Omega}{d \ln R} \right|^2 \nu \Sigma \Omega^2 \approx \left| \frac{d \ln \Omega}{d \ln R} \right| \frac{\dot{M}}{2\pi} \Omega^2,$$

where the last equality holds approximately at large radii, where we can use Eq. (48). The dissipation rate is therefore proportional to  $\Omega^2$ , which can be a relatively steep decreasing function of radius. For example, if the potential is dominated by the central point mass and the rotation curve is Keplerian, we have  $D(R) \propto R^{-3}$ . In the other extreme condition of a purely self-gravitating disc, with a flat rotation curve, the decrease is slightly more gentle, with  $D(R) \propto R^{-2}$ . In fact, whenever the disc contributes to the rotation curve, it results in a gentler decrease of  $\Omega$  at large radii, therefore producing a relatively larger energy dissipation, especially in the outer disc.

The above discussion provides a way to estimate to first approximation what is the spectral energy distribution (SED) produced by an actively accreting disc. Let us assume that the energy dissipated through viscosity is released as blackbody emission at the disc surface (which would occur if the disc is sufficiently opaque). Then the effective temperature  $T_{\text{eff}}$  of the emitted spectrum is simply given by  $2\sigma_B T_{\text{eff}}^4$ , where  $\sigma_B$  is Stefan-Boltzmann constant and where the factor 2 is introduced because the energy is emitted by the two sides of the disc. This then leads to

$$(51) \quad T_{\text{eff}}(R) = \left( \frac{\dot{M}}{4\pi\sigma_{\text{SB}}} \left| \frac{d \ln \Omega}{d \ln R} \right| \Omega^2 \right)^{1/4} = \left( \frac{3GM\dot{M}}{8\pi\sigma_{\text{SB}}R^3} \right)^{1/4},$$

where the last equality holds for a Keplerian rotation curve. Each annulus of the disc thus emits as a blackbody at a different temperature. The full spectrum of the disc is then obtained by superimposing the different blackbody spectra from different annuli. If  $F_\lambda$  is the flux emitted at wavelength  $\lambda$ , we thus have

$$(52) \quad F_\lambda = \frac{\cos i}{d^2} \int_{R_{\text{in}}}^{R_{\text{out}}} 2\pi R B_\lambda[T_s(R)] dR,$$

where  $d$  is the distance between the disc and the observer,  $i$  is the inclination and  $B_\lambda(T)$  is the Planck function. The shape of the SED is a ‘stretched’ blackbody spectrum, extending over a much wider wavelength range than a simple Planck spectrum, from the



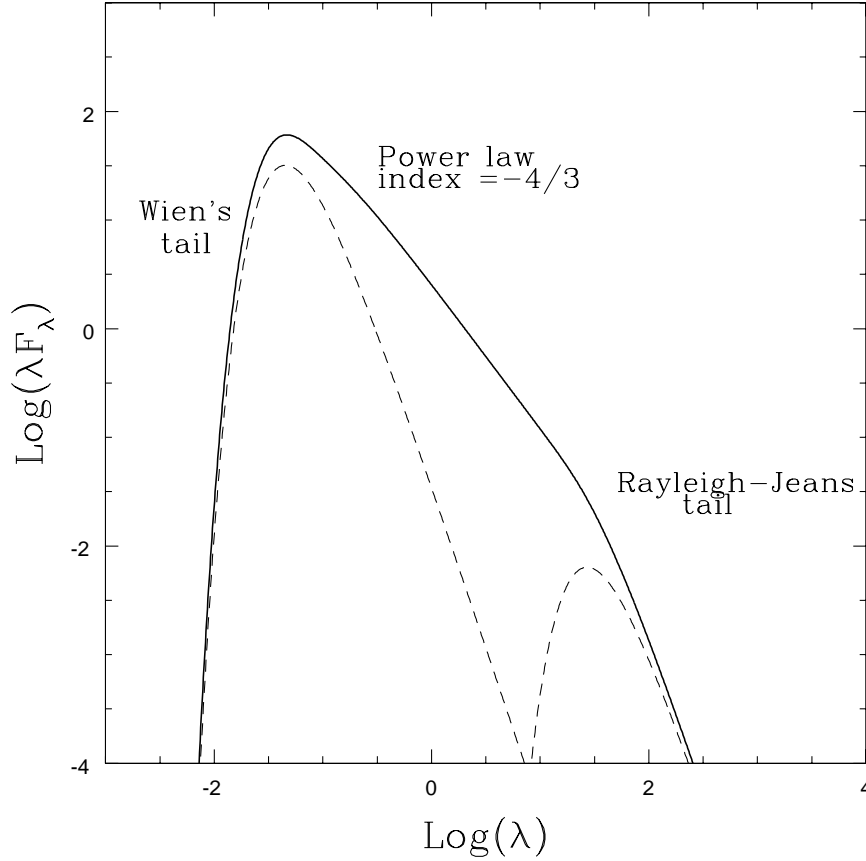


Fig. 3. – SED of a Keplerian optically thick disc. At short wavelength the SED approaches the Wien's tail corresponding to the temperature in the inner disc  $T_{\text{in}}$ . At large wavelengths the SED approaches the Rayleigh-Jeans tail corresponding to the outer disc temperature  $T_{\text{out}}$ . At intermediate wavelengths the SED is a power law with index depending on the temperature scaling with radius. For the Keplerian disc shown here the power law index is  $n = -4/3$ . The two dashed lines show for comparison two pure blackbody spectra corresponding to  $T_{\text{out}}$  and  $T_{\text{in}}$ . Units are arbitrary.

shortest wavelength  $\lambda_{\text{min}} \approx hc/kT_{\text{in}}$ , associated to the temperature in the inner disc, to the longest  $\lambda_{\text{max}} \approx hc/kT_{\text{out}}$ , corresponding to the outer disc temperature. Fig. 3 shows a typical SED produced by a Keplerian disc, with the two dashed lines corresponding to blackbody spectra at  $T = T_{\text{min}}$  and  $T = T_{\text{max}}$ . The short wavelength part of the SED approaches the Wien's tail appropriate to the inner disc temperature  $T_{\text{in}}$ . The long wavelength SED approaches the Rayleigh-Jeans tail appropriate to the outer disc temperature  $T_{\text{out}}$ . At intermediate wavelengths, it can be shown [37, 38] that, if the temperature profile is a power law with respect to radius, with power law index  $-q$ , then the SED  $\lambda F_{\lambda}$  is also a power law with respect to wavelength, with power law index  $n = 2/q - 4$ . We thus get the typical Keplerian accretion disc spectrum, where  $q = 3/4$  and thus  $n = -4/3$ . For a flat rotation curve disc, we have  $q = 1/2$ , which then leads

to  $n = 0$  and the SED is flat at intermediate wavelengths. This reflects the fact that for a flatter rotation curve the emission at large distances is enhanced, thus enhancing the long-wavelength contribution to the SED.

It is interesting to notice that in many cases the observed SED from accretion discs (for example, in the case of protostellar discs) is indeed relatively flat, indicating a flat temperature profile. However, in most cases, the disc mass is not high enough to produce substantial deviations from Keplerian rotation in order to reproduce such flat SED, except possibly for some interesting cases (and in particular for the so called FU Orionis objects), which will be discussed in more detail in Section 7.1 below. The observational evidence for a relatively hot outer disc is generally interpreted in terms of an additional source of heating, coming for example from irradiation from the central star or from the inner disc. A full treatment of the energy balance in an accretion disc requires solving the radiative transfer equation through the disc, and therefore depends on the detailed optical properties of the gas. This has been done under a variety of conditions in several papers [39, 40].

Finally it is worth pointing out the relationship that is established between the viscosity parameter and the cooling time in thermal equilibrium. Indeed, in equilibrium (and assuming that there is no other source of heating for the disc) the heating rate provided by viscosity has to balance the cooling rate. The cooling rate per unit surface can be simply written as

$$(53) \quad C = \frac{\Sigma c_s^2}{\gamma(\gamma - 1)\tau_{\text{cool}}},$$

where  $\gamma$  is the ratio of specific heats and where we have introduced a cooling timescale  $\tau_{\text{cool}}$ . Equating this expression to the heating rate due to viscosity (Eq. (50)) and using the  $\alpha$ -prescription we get

$$(54) \quad \alpha = \left| \frac{d \ln \Omega}{d \ln R} \right|^{-2} \frac{1}{\gamma(\gamma - 1)\Omega\tau_{\text{cool}}}.$$

### 3. – GRAVITATIONAL INSTABILITIES AND SELF-REGULATION

In the sections above, we have described the equilibrium configuration of accretion discs and their secular evolution due to viscosity. We have described what is the effect of the disc self-gravity on such equilibrium state. In particular, we have seen that for small but non negligible total disc masses  $M_{\text{disc}}/M \approx H/R$ , self-gravity affects the disc vertical structure, while the rotation curve is only affected by self-gravity when the disc mass becomes comparable with  $M$ . However, one of the most important effects associated with self-gravity is the possibility of developing gravitational instabilities. The physical origin of the instability is essentially related to the standard Jeans instability for a homogeneous fluid, where large scale disturbances cannot be stabilized by pressure gradients, thus determining a typical wavelength (the so-called Jeans wavelength) associated with the instability. In a rotating disc, as I will show here below, the additional effect of rotation can also stabilize these large scale perturbations.

As I have already remarked above, gravitational instabilities might be particularly important in providing a significant source of transport of angular momentum through the disc. Under certain conditions, the non-linear outcome of the instability can be relatively violent, leading to the fragmentation of the disc into gravitationally bound clumps.

Most of the recent research on self-gravitating discs has concentrated on exploring the conditions under which the instability takes the form of a long-lived spiral structure, leading to a steady transport of angular momentum, or alternatively determines the disc fragmentation. In this section and in the following ones, I discuss in detail such issues.

**3.1. Development of gravitational instabilities.** – The dynamical processes underlying the development and the maintenance of gravitational instabilities in discs have been discussed in several papers and reviews, mostly in view of an application to the discs of spiral galaxies [12, 25]. On the other hand, many processes relevant to the dynamics of stellar discs are also applicable to the case of a fluid disc. The starting point of such analysis is the determination of the propagation properties of density waves in a self-gravitating disc, through a standard linear perturbation analysis of the relevant dynamical equations, i.e. the continuity equation (Eq. (2)), the momentum equation for an inviscid disc (Eq. (3)), supplemented by the equation of state (Eq. (7)), to relate pressure and density, and by Poisson’s equation (Eq. (4)) in order to connect the disc density to the gravitational potential. The difficulty here arises from the long-range nature of gravitational force, so that the gravitational potential at a given location depends on the whole distribution of the gas density  $\Sigma$ , rather than on its local value. A local analysis can nonetheless be carried out in the WKB approximation, for tightly wound perturbations, for which the pitch angle  $m/(kR) \ll 1$ , where  $k$  is the radial wavenumber and  $m$  is the azimuthal wavenumber (corresponding to the number of arms of a spiral disturbance). In this approximation, the dispersion relation is:

$$(55) \quad (\omega - m\Omega)^2 = c_s^2 k^2 - 2\pi G \Sigma |k| + \kappa^2,$$

where  $\omega$  is the frequency of the perturbation and the epicyclic frequency  $\kappa$  is given by

$$(56) \quad \kappa^2 = \frac{2\Omega}{R} \frac{d(\Omega R^2)}{dR}.$$

Note that the epicyclic frequency  $\kappa$  is generally of the order of the angular velocity  $\Omega$ , the exact proportionality depending on the rotation curve. For example, for Keplerian rotation, we have  $\kappa = \Omega$ , while for a flat rotation curve, we have  $\kappa = \sqrt{2}\Omega$ .

A detailed discussion of the physics behind this dispersion relation can be found elsewhere [1, 25]. For our purposes it suffices to describe the most salient features of it. Let us consider axisymmetric disturbances, for which  $m = 0$ . Clearly, whenever  $\omega^2$  is positive, a given perturbation simply propagates in a wave-like way. On the other hand, if  $\omega^2$  is negative, an exponentially growing instability arises. We thus see from the right-hand side of Eq. (55) that the second term, which is negative and is associated with the disc self-gravity, is a potentially destabilizing term, while the first, associated with pressure forces, and the third, related to rotation, are stabilizing terms. In particular, while pressure stabilizes the disc at small wavelengths (large  $k$ ), rotation is more effective at stabilizing large scale disturbances (small  $k$ ). Equation (55) is a simple quadratic expression for the wavelength  $k$  and it can be easily seen that for axisymmetric perturbations the right hand side is positive definite (i.e.  $\omega^2$  is positive and the perturbation is stable) if:

$$(57) \quad Q = \frac{c_s \kappa}{\pi G \Sigma} > 1.$$

The dimensionless parameter  $Q$  is therefore essential in determining the local axisymmetric stability of self-gravitating accretion discs (a similar parameter also determines the

stability properties of stellar discs, [24]). We have already encountered a parameter very similar to  $Q$ , when discussing the relative importance of the central object and of the disc in determining the disc thickness (Eq. 27). We had seen that when  $c_s\Omega/\pi G\Sigma \approx 1$  then the disc self-gravity contributes to the vertical hydrostatic balance equation as much as the central object. We now see that for the very same range of parameters we also expect the disc to be close to marginal gravitational instability. Note that in the tight winding approximation  $m$  does not enter explicitly in the dispersion relation (except in the Doppler shifted frequency) so that both axisymmetric perturbations ( $m = 0$ ) and non-axisymmetric ones ( $m \neq 0$ ) share the same instability properties.

For discs that are gravitationally unstable, that is in cases where  $Q \lesssim 1$  in the equilibrium state, the most unstable wavelength can be easily obtained by finding the wavenumber  $k_0$  for which the right-hand side of Eq. (55) attains a minimum. This is given by

$$(58) \quad k_0 = \frac{\pi G \Sigma}{c_s^2} = \frac{1}{H_{\text{sg}}}.$$

We thus see that the most unstable wavelength is of the order of the disc thickness. The condition for the WKB approximation to hold is then  $kR = R/H_{\text{sg}} \gg 1$  and it is satisfied for marginally stable discs if the disc is thin, or equivalently if the disc mass is much smaller than the central object mass,  $M_{\text{disc}} \ll M$  (cf. Eq. (26)).

It is well known (for example, from early  $N$ -body numerical simulations [41, 42]) that massive discs are subject to violent large scale (long-wavelength) non-axisymmetric instabilities, even in cases where the parameter  $Q$  is above unity, and therefore the disc is predicted to be stable in the WKB approximation. Indeed, a local dispersion relation can also be obtained in the case of relatively long radial wavelengths in the non-axisymmetric case [43]. This relation is more complicated than the simple quadratic form described above and depends on a new parameter, in addition to  $Q$ , defined as:

$$(59) \quad J = m \frac{\pi G \Sigma}{R \kappa^2} \frac{4\Omega}{\kappa} \left| \frac{d \ln \Omega}{d \ln R} \right|^{1/2}.$$

Being proportional to  $\Sigma$ ,  $J$  is a measure of how massive the disc is. Indeed, to give an idea of the numbers involved, consider a quasi-Keplerian disc, where  $\kappa \approx \Omega$  and where  $\Sigma \propto R^{-1}$ . In this case we have  $J = \sqrt{6} m M_{\text{disc}}(R)/M$ . If  $J \ll 1$  (light disc or zero  $m$ ), the dispersion relation reduces to the standard quadratic form. On the other hand, for  $J \approx 1$ , the perturbations are significantly more unstable. A detailed description of this regime can be found elsewhere [25, 44]. In particular, when the disc mass is a sizable fraction of that of the central object, low  $m$  disturbances can be violently unstable. Interestingly, as I will discuss below, this change of behaviour between light and massive discs has been observed numerically in a recent survey of the development of gravitational instabilities in fluid discs [45, 46].

To summarize, in the tightly wound approximation we expect a gaseous disc to become gravitationally unstable when  $Q < 1$ . Non-axisymmetric disturbances, with large azimuthal wavenumber  $m$  can be unstable even for relatively larger value of  $Q$ , even though if  $Q$  rises above  $\approx 2$ , then even non-axisymmetric disturbances are stabilized. Finally, it is worth noting that taking into account finite-thickness effects in the tightly wound approximation leads to a stabilization of the disc, so that the marginal stability value of  $Q$  is reduced below unity [47].

**3.2. Self-regulation.** – Having established that massive discs are linearly unstable, we should now ask what is the non-linear evolution of the instability and whether it can lead to a sustained transport of angular momentum. The stability criterion, Eq. (57) offers a natural way to describe this. Indeed, we should note that the stability parameter  $Q$  is proportional to the sound speed (and hence to temperature), so that colder discs are more unstable. Now, let us consider a disc which is initially hot, so that  $Q \gg 1$  and the disc is stable. In the absence of other transport and heating mechanisms, the disc cools down due to radiative cooling until eventually  $Q \approx 1$ . At this stage, the disc develops a gravitational instability in the form of a spiral structure. Compression and shocks induced by the instability lead to an efficient energy dissipation and as a result the disc will heat up. In turn, the stability condition works like a kind of ‘thermostat’, so that heating turns on only if the disc is colder than a given temperature. If the ‘thermostat’ works, we should then expect the instability to self-regulate in such a way that the disc is always kept close to marginal stability, and thus a self-gravitating disc will evolve to a state where  $Q \approx \bar{Q}$ , where  $\bar{Q}$  is a constant of order unity. Here again, it is reassuring to see that numerical simulations [45] reproduce closely this behaviour.

Note here an important difference between a stellar and a fluid disc. In order for the above self-regulation process to work, the presence of a dissipative component is essential. While this is obviously the case for a gaseous disc, a pure stellar disc lacks this dissipation channel. For a galaxy disc, therefore, self-regulation requires the presence of a gaseous component in addition to the dominant stellar one [48].

**3.3. Self-regulated accretion discs models.** – The self-regulation mechanism described above can be easily introduced in steady state models of self-gravitating accretion discs [15]. Indeed, for a self-gravitating disc, as shown in Section 2.7 at large radii we have:

$$(60) \quad \dot{M} = \frac{2\alpha c_s^3}{G} \left| \frac{d \ln \Omega}{d \ln R} \right|,$$

where we have also used the  $\alpha$ -prescription. If we supplement this equation by the self-regulation condition:

$$(61) \quad Q = \frac{c_s \kappa}{\pi G \Sigma} = \bar{Q} \approx 1,$$

we thus easily see that in the case of a disc dominated potential (that is, if the central star mass is negligible) a Mestel disc with  $\Sigma \propto R^{-1}$  provides an analytical self-similar solution to the model. In particular, we have:

$$(62) \quad c_s = \left( \frac{G \dot{M}}{2\alpha} \right)^{1/3},$$

$$(63) \quad v_c = \frac{2\sqrt{2}}{\bar{Q}} \left( \frac{G \dot{M}}{2\alpha} \right)^{1/3},$$

$$(64) \quad 2\pi G \Sigma R = \frac{8}{\bar{Q}^2} \left( \frac{G \dot{M}}{2\alpha} \right)^{2/3}.$$

The solution above has remarkably simple properties, with a flat profile of  $v_c$ ,  $c_s$  and  $Q$  and where the surface density has a simple power-law dependence on radius. However, it is clearly approximate and only applies to large radii, where the effects of  $\dot{J}$  and especially of the presence of a central point mass  $M$  can be neglected. In particular, in realistic accretion discs the latter condition is difficult to achieve and we should therefore extend the analysis to include a central mass in the model. In particular, we expect the transition from an inner Keplerian disc to an outer disc, where the rotation curve would tend to flatten, to occur at a distance of the order of

$$(65) \quad R_s = 2GM \left( \frac{\bar{Q}}{4} \right)^2 \left( \frac{GM}{2\alpha} \right)^{-2/3}.$$

In this way we can use the simple model above to check the influence of the disc self-gravity on the rotation curve of some observed systems. Let us then estimate the transition radius  $R_s$  for a number of different cases. We first consider a typical protostellar disc, for which  $\dot{M} \approx 10^{-8} M_\odot/\text{yr}$ , the central star mass is of the order of  $1 M_\odot$  and where we assume  $\alpha = 0.01$ . Inserting these numbers in Eq. (65), we get  $R_s \approx 7000 \text{ AU}$ . We thus see that in order for self-gravity to be important, the disc should extend out to very large distances, much larger than the observed protostellar disc sizes. Typical protostellar discs should therefore not be significantly non-Keplerian. On the other hand, as already mentioned above, in some cases the accretion rate of young protostellar discs can be significantly larger. For example, during the so-called FU Orionis outbursts (see below Section 7.1 for more details) the accretion rate can reach  $\dot{M} \approx 10^{-4} M_\odot/\text{yr}$ . In such cases,  $R_s$  is reduced to  $\approx 10 - 20 \text{ AU}$ , so that the outer parts of the disc might show some deviation from Keplerian rotation. Finally, let us consider a typical disc in an Active Galactic Nucleus, for which  $\dot{M} \approx 0.1 M_\odot/\text{yr}$  and  $M \approx 10^7 M_\odot$ . In this case, we get  $R_s \approx 5 \text{ pc}$ . Interestingly, as mentioned in the Introduction, water maser emission from AGN discs has been observed exactly at such distances and we therefore expect that at least some of them will display some degree of non-Keplerian rotation.

In order to compare these self-regulated models to observations, it is then important to extend the analytical self-similar solution to the more general case where it connects to an inner Keplerian disc. Such models have been constructed [13] and some examples of rotation curves obtained in this more general case can be seen in Fig. 4. Such models can be used for a detailed comparison with observed rotation curves, as will be discussed in Section 8 below. Additionally, it is worth mentioning that numerical simulations which try to reproduce the formation of protostellar discs [50], tend to be characterized by fairly massive discs in their earliest stages, with properties that resemble closely the ones described semi-analytically by these models.

Self-regulated disc models also tend to be relatively hotter in their outer regions, compared to their non-self-gravitating counterparts. This is due both to the self-regulation mechanism, and the effective heating associated with it, and by the fact that the disc has a flatter rotation curve in its outer parts, as discussed in Section 2.8. As a consequence, self-regulated SEDs tend to be flatter than the standard disc SED, shown in Fig. 3. and often display a secondary “bump” at long wavelengths [49]. An example SED is shown in Fig. 4, where the different curves refer to different choices of the disc outer radius (see [49] for details). These models were originally proposed in order to explain some observations of protostellar discs, and will be discussed more in detail in section 7.1 below. Interestingly, very similar SEDs are also produced in numerical sim-

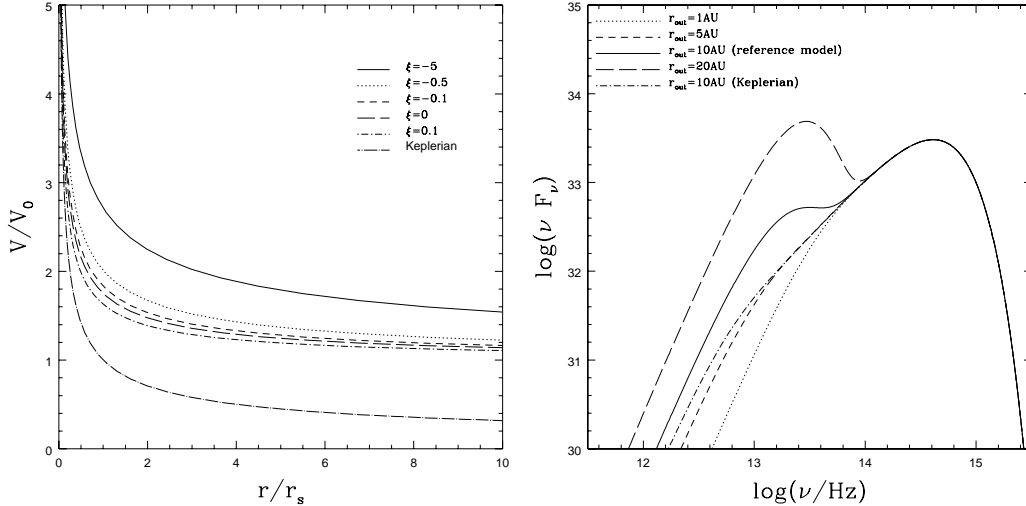


Fig. 4. – Left: examples of rotation curves for self-regulated disc models in the presence of a central point mass. In the inner region the rotation curve is Keplerian, while it becomes progressively flatter in the outer disc, asymptotically reaching the self-similar configuration. More details can be found in [13]. Right: example SEDs for self-regulated discs. At large radii self-regulated discs are hotter than their non-self-gravitating counterparts, resulting in a secondary bump in the SED at long wavelengths [49], which becomes more pronounced if the disc outer radius is increased.

ulations of self-gravitating protostellar discs [51], and similar semi-analytic models for self-gravitating disc SEDs have later been proposed also for AGN discs [52].

#### 4. – NUMERICAL SIMULATIONS OF SELF-GRAVITATING ACCRETION DISCS

In the sections above we have discussed the development of gravitational instabilities essentially in terms of linear perturbation analysis. We have already noted that the non-linear evolution of such instabilities can result in a feedback loop which self-regulates the disc and keeps it close to marginal stability. It is then possible, as we have seen, to develop analytical models where such complex behaviour is introduced in a simple way in the basic steady state disc equations. Clearly, in order to check whether self-regulation is indeed established once the instability becomes non-linear, and to investigate secular effects, like the transport of angular momentum induced by the spiral structure, it is important to perform numerical simulations of the disc well into the non-linear regime. Many simulations of this kind were performed in the early '70s in the context of collisionless stellar systems by means of  $N$ -body techniques [41, 42, 53, 54].

Here, we are interested in the description of a dissipative, gaseous disc and thus these early  $N$ -body simulations were not able to catch some aspects of the disc hydrodynamics. In the last fifteen years complex, three-dimensional hydrodynamic simulations have become an essential tool to study accretion discs dynamics and have provided an important ‘experimental’ counterpart to analytical investigations, such as those described in the previous sections. In this respect, a set of simulations should be considered as a ‘nu-

merical experiment’ where our models are checked against measurements of a controlled system, often unavailable from astronomical observations. Simulations have evolved significantly throughout the years, partly due to the increased computational power (which has enable to reach higher resolution and accuracy), and partly to the development of progressively more refined algorithms to include new physical processes in the models.

A key role in the evolution of self-gravitating accretion discs is played by the cooling process. As discussed above, it is the interplay between cooling and internal heating due to gravitational instability that allows to reach a self-regulated condition. As will be discussed in more detail in section 5 below, the cooling rate also determines the efficiency with which the gravitational instability redistributes angular momentum and allows accretion. Most of the effort in this area of research is therefore aimed at introducing into the codes progressively more realistic implementations of the cooling processes in accretion discs. In this section we provide an overview of the development of numerical simulations of accretion discs in the last few years. Specific results concerning the transport induced by gravitational instability and the possibility of disc fragmentation are discussed in Sections 5 and 6 below.

Some early attempts at simulating self-gravitating accretion discs were made by using collisionless  $N$ -body methods [55]. However, the earliest hydrodynamical simulations of self-gravitating accretion discs were performed later [56] in the context of the formation of protostellar discs. Such simulations employed a very simplified thermodynamical description of the gas, that was considered to be ‘locally isothermal’ in the sense that every parcel of fluid was assumed to keep its internal energy throughout the simulation. Several other early simulations were performed using different equations of state, such as a polytropic equation [57,58] and were used to describe the development of gravitational instability, but the effects of the choice of thermal state of the gas were generally overlooked. More emphasis on thermodynamical aspects was placed later [59,60], when a comparison between models evolved isentropically and models which enforced an isothermal behaviour revealed that gravitational instabilities developed much more violently in an isothermal configuration, leading to a very fast redistribution of angular momentum and in extreme cases to the disruption of the disc. A posteriori, this behaviour is simply understood based on the arguments described in Section 3.3 above. The assumption of isothermality prevents the development of the instability to feed back thermal energy in the equilibrium state, therefore not allowing the non-linear stabilization of the disturbance and the saturation of the gravitationally unstable modes.

All the simulations described so far start from an initial configuration which is expected to be highly unstable, with  $Q$  values of the order of unity or below. It is therefore not surprising that as soon as the simulation is set to evolve, strong disturbances suddenly appear and grow out of control in the case of isothermal simulations. The next question is therefore what would be the response of a disc which is initially stable and is driven to an unstable configuration through cooling. Clearly, in order to consider this case, one has to move away from either an isothermal or a polytropic equation of state and consider instead the full evolution of the energy equation. This was initially done by considering a simple cooling prescription [45,46,61-68] where the cooling rate is given by:

$$(66) \quad \left. \frac{du}{dt} \right|_{\text{cool}} = - \frac{u}{\tau_{\text{cool}}},$$

where  $u$  is the internal energy of the fluid and  $\tau_{\text{cool}}$  is the cooling timescale, which is set as



a free parameter in the simulations. It is important to stress that the above description is not meant to reproduce any specific cooling law, but is just a convenient way (a ‘toy’ model) of exploring the role of the cooling timescale in the outcome of the gravitational instability. Different investigators use different prescriptions for the specific form of the cooling time, which in some cases is taken to be a constant while in other cases is taken to be proportional to the dynamical timescale, so that  $\Omega\tau_{\text{cool}}$  is kept constant. Despite the different prescriptions for the cooling time and the different type of codes used (local shearing sheet models [61], global Eulerian grid-based models [66, 67], global Lagrangian particle-based models [45, 46, 62–65, 68]) the results appear to converge into an essentially coherent picture, at least for low mass discs. It is now well established that the long term behaviour depends on the relative values of the cooling and dynamical timescale. If the cooling timescale is larger than a few dynamical timescales, an initially stable (large  $Q$ ) disc cools down until  $Q$  becomes of the order of unity. At this stage, the disc becomes gravitationally unstable and develops a spiral structure which provides a heating source, through compressional heating and shock dissipation, able to balance the externally imposed cooling. Once in thermal equilibrium, the disc is characterized by an approximately constant value of  $Q$  very close to marginal stability. In such a state a spiral structure persists in the disc, to provide the required heating. Therefore the self-regulation mechanism described above determines the disc structure and evolution. Figure 5 shows the result of one such simulations, where in this case  $\tau_{\text{cool}}\Omega = 7.5$  and the total disc mass  $M_{\text{disc}} = 0.1M$  [45]. In the left panel the colour plot shows the disc surface density, in which a spiral structure is clearly seen. The right panel shows the azimuthally and vertically averaged value of  $Q$  as a function of radius. The disc in this case extends from  $R = 0.25$  to  $R = 25$  in code units. It is then seen that far from the boundaries (where the density drops and  $Q$  correspondingly grows) the disc is self-regulated, with  $Q \approx 1$  over a wide radial range. The three curves refer to three different times during the evolution, showing that the disc has reached thermal equilibrium.

In [45] several simulations are performed, with varying mass ratio between the disc and the star,  $M_{\text{disc}}/M = 0.05, 0.1$  and  $0.25$ . This allows a comparison of the morphology and of the dynamics of the spiral structure developed in the different cases. The upper panel of Fig. 6 shows a snapshot of the density structure three simulations once they have reached a self-regulated state, while the lower panel shows the power associated with Fourier components of different  $m$ . Several interesting features can be noted. First of all, note that the typical radial wavelength of the instability increases as the disc mass increases. This can be easily understood even from the simple quadratic dispersion relation, Eq. (55), which predicts that the typical wavelength of the instability is proportional to the disc thickness  $H$ , Eq. (58). Now, as we know, self-regulated discs have  $H/R \approx M_{\text{disc}}/M$  and therefore a more massive disc is also thicker and the resulting spiral structure tends to be more open. Also, note that while for the low mass case the different Fourier modes have similar amplitudes, as the disc mass increases, modes with lower  $m$  tend to become more important. Also this behaviour can be understood in terms of the local dispersion relations described above. Indeed, when the disc mass is low, the parameter  $J \approx mM_{\text{disc}}/M$  is also small (except for very large  $m$  which are not easily excited) and we can use the quadratic dispersion relation which predicts a similar behaviour for all  $m$ . On the other hand, for larger disc masses,  $J$  can be of order unity already for small  $m$  which then tend to be much more unstable, as confirmed in the simulations. If we increase the mass ratio further, going up to  $M_{\text{disc}} \sim M$  [46] this trend continues and strong  $m = 2$  disturbances become violently unstable. In this high mass case, self-regulation cannot be established and the disc shows a recurrent pattern of highly

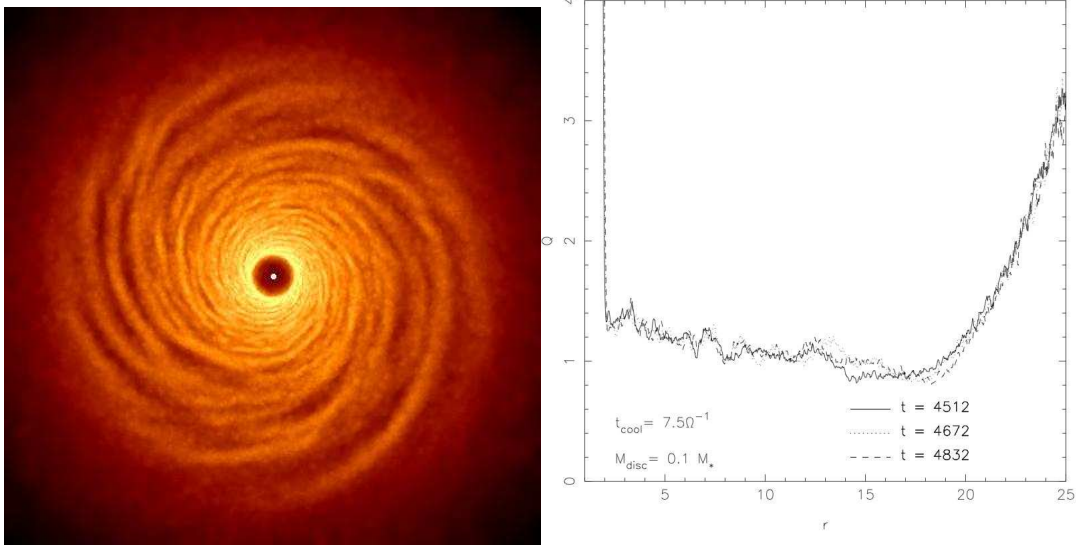


Fig. 5. – Numerical simulation of the development of a gravitational instability in a cooled disc with  $\tau_{\text{cool}}\Omega = 7.5$  and where the disc mass  $M_{\text{disc}} = 0.1M$  [45]. Left: Surface density of the disc, showing the prominent spiral structure. Right: radial profile of the azimuthal and vertical average of  $Q$ . The disc is self-regulated over a wide radial range. The three curves refer to three different times during the evolution, showing that the disc has reached thermal equilibrium.

temporally variable instabilities, as shown in Fig. 7. During such episodes, the torques induced by self-gravity can be very large, and cause a relatively fast redistribution of gas within the disc.

The behaviour described above changes when the cooling time is decreased to smaller values [61]. In this case, the disc does not reach a quasi-steady self-regulated state but rather fragments into several bound objects. Figure 8 show the results of a simulation very similar to the one displayed in Fig. 5, but where the cooling time is decreased to  $\tau_{\text{cool}} = 3\Omega^{-1}$  [65]. This result can be understood in the following way, by adopting a local approach to describe the instability. In a gravitationally unstable disc, the typical growth timescale of unstable perturbations is of the order of the dynamical timescale  $\Omega^{-1}$ . The non-linear stabilization of the perturbation only works if the heat generated by compression and shocks is not removed efficiently from the disc through cooling. Since the perturbation grows on the dynamical timescale, if we want to avoid fragmentation, we require that cooling acts on a longer timescale. Note that the requirement that the cooling timescale be shorter than the dynamical timescale in order to result in fragmentation has been known for several years, even outside the context of disc instability [69, 70]. The exact value of the threshold for fragmentation does depend slightly on the specific numerical set-up and ranges from  $\tau_{\text{cool}} = 3\Omega^{-1}$  to  $\tau_{\text{cool}} = 6\Omega^{-1}$  [61, 65, 71]. An additional discussion on the dependence of the threshold on the ratio of specific heats and on its interpretation is provided in Section 6 below.

More recent simulations have further improved the thermodynamical description of the disc, including a more realistic cooling prescription based on the opacity properties of the gas. In the simplest case, one can implement such ‘realistic’ cooling in local

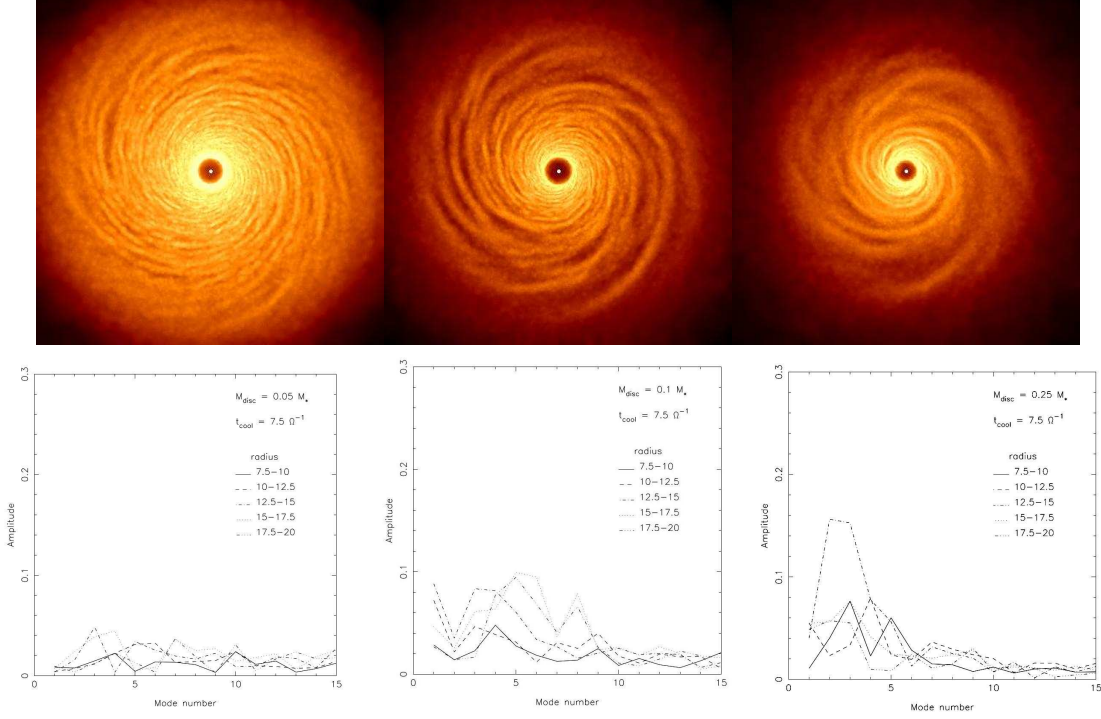


Fig. 6. – Top panels: Surface density structure for a set of self-gravitating disc simulations with  $\tau_{\text{cool}}\Omega = 7.5$ . From left to right the mass ratio is  $M_{\text{disc}}/M = 0.05, 0.1$  and  $0.25$ . Lower panels: corresponding mode analysis. The amplitude of various modes with different  $m$  is shown.

simulations of infinitesimally thin discs, where the vertical radiative transfer is simply treated with a one-zone approximation [72]. More realistic, three-dimensional simulations including radiative transfer are being developed in the last few years [73] and start to provide their first results [51, 74, 75].

## 5. – TRANSPORT INDUCED BY GRAVITATIONAL INSTABILITIES

The discussion in the last two sessions has shown that (a) accretion discs can be subject to gravitational instabilities whenever the parameter  $Q$  (which describes linear stability in the tightly wound approximation) is of order unity and (b) under appropriate conditions the non-linear evolution of the instability tends to self-regulate in such a way that the disc is kept very close to marginal stability ( $Q \approx 1$ ) over a wide radial range. We have also seen that, in the limit of Keplerian rotation, the condition  $Q \approx 1$  is roughly equivalent to  $M_{\text{disc}}/M \approx H/R$ , so that if the thickness is small, even a relatively light disc can be gravitationally unstable.

We now turn to the question of the transport of energy and of angular momentum induced by the instability. In a standard viscous accretion disc, both energy and angular momentum are transported through the disc by viscosity, and are therefore described by a single transport coefficient. As we will see, this might not be the case for self-gravitating accretion discs, and we will then have to determine if and under which conditions can

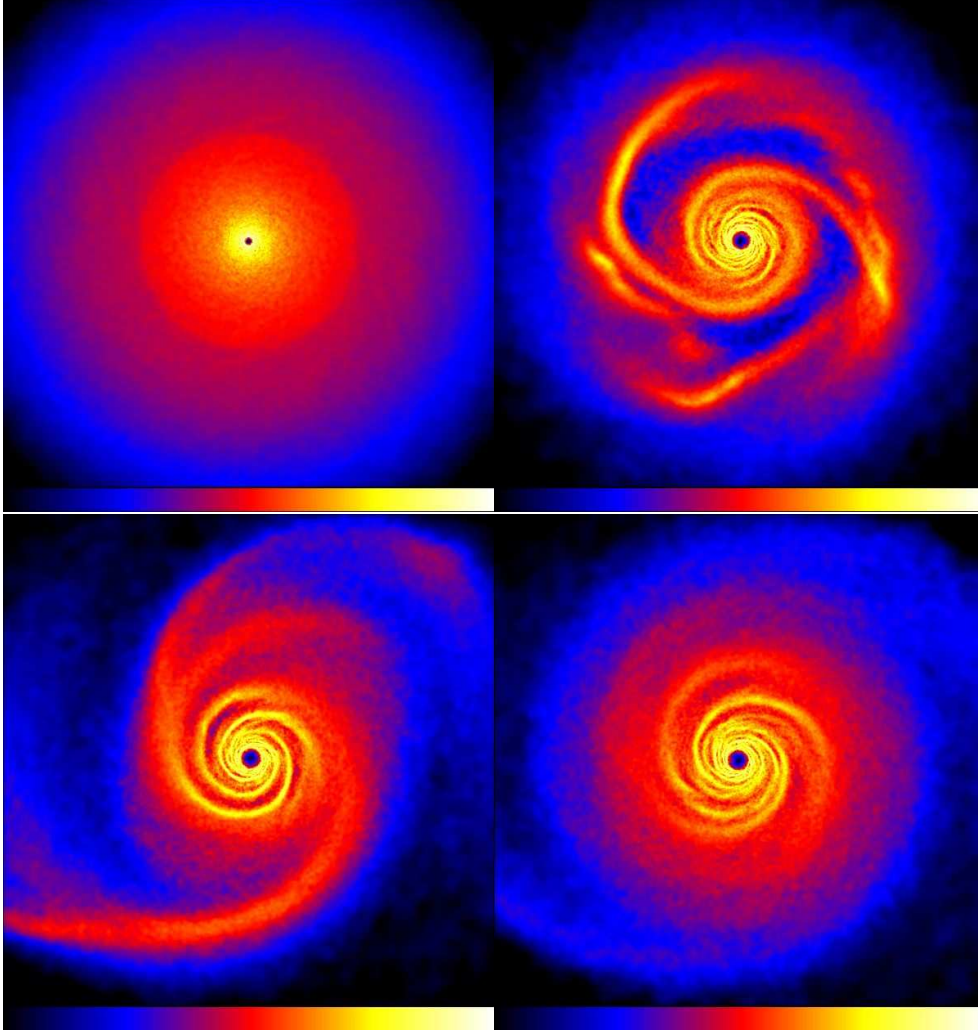


Fig. 7. – Evolution of the surface density of the disc during the development of a transient episode of spiral activity in a simulation of a massive  $M_{\text{disc}} = 0.5M$  accretion disc. The four snapshots refer to four different times during the simulation,  $t = 0$  (upper left),  $t = 1.9$  (upper right),  $t = 2.5$  (lower left),  $t = 3$  (lower right), where the times are given in units of the orbital period at the disc outer edge. From [46].

such a self-gravitating disc be described in terms of a single ‘viscosity’ coefficient. That a spiral density wave transports angular momentum was early recognized in the context of stellar dynamics [76, 77]. In the context of accretion discs, where most of the theoretical development has been achieved through the use of an *ad hoc* (cf. the  $\alpha$ -prescription) anomalous viscosity, the key question is understanding to what extent can this form of transport be described in terms of a ‘viscous’ process. In other words, the  $\alpha$ -prescription for viscosity has always been used with the (often tacit) assumption that it actually corresponds to the transport induced by some kind of disc instability. It is therefore

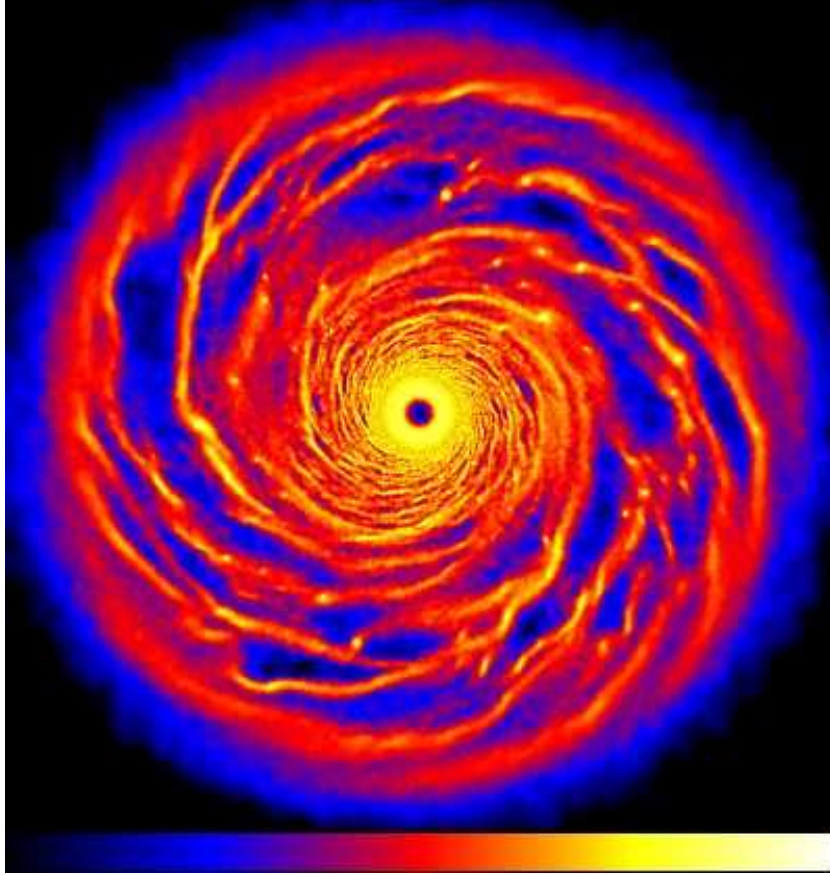


Fig. 8. – Numerical simulation of a self-gravitating disc with  $M_{\text{disc}} = 0.1M$  and  $\tau_{\text{cool}} = 3\Omega^{-1}$ . Once unstable, the disc breaks up into numerous gravitationally bound clumps.

natural to ask whether the development of gravitational instability can indeed be this long sought cause of angular momentum transport.

A possible way to proceed is to simply *assume* that the transport of angular momentum takes the same form of a viscous diffusion and to introduce a  $Q$ -dependent viscosity parameter  $\alpha$  [9,10], such that

$$(67) \quad \alpha_{\text{sg}} = \begin{cases} \xi \left( \frac{\bar{Q}^2}{Q^2} - 1 \right) & Q < \bar{Q} \\ 0 & Q > \bar{Q} \end{cases}$$

Here  $\bar{Q}$  is the value of  $Q$  at which the disc becomes unstable to non axisymmetric perturbations and  $\xi$  is a parameter to measure the strength of the induced torques. The above formulation is useful in practical cases, for example, when one wants to incorporate in a simple way the self-regulation mechanism in simple time-dependent models of self-gravitating discs. However, it lacks one important feature elucidated from the nu-



merical simulations described above. In this picture,  $\alpha_{\text{sg}}$  only depends on the local value of  $Q$  and not on the cooling timescale  $\tau_{\text{cool}}$ , that we have seen controls so efficiently the development of the instability. In particular, for self-regulated discs, we expect  $Q \approx \bar{Q}$  and the formula above would then produce a negligibly small  $\alpha_{\text{sg}}$ , while we know that a finite amplitude spiral structure is present in self-regulated discs and indeed it is this spiral structure that provides the heating to balance the imposed cooling rate.

If we maintain the assumption that the transport induced by self-gravity can be described in terms of viscosity, a prescription for  $\alpha_{\text{sg}}$  consistent with the apparent behaviour of the numerical simulations described above can be obtained by the following argument, which is inspired from the self-regulation mechanism. We have seen that the behaviour of the disc is strongly dependent on the cooling timescale. In particular, for a given cooling timescale, the disc eventually reaches a steady thermal equilibrium state, where the amplitude of the spiral instability saturates at a value such that heating balances the imposed cooling. If we now decrease the cooling time, the disc will cool down and become more unstable, and the saturation amplitude of the instability will grow, thus providing an enhanced transport and dissipation, to match the new cooling configuration. A simple prescription can then be provided from the requirement of thermal equilibrium (cf. Eq. (54))

$$(68) \quad \alpha_{\text{sg}} = \left| \frac{d \ln \Omega}{d \ln R} \right|^{-2} \frac{1}{\gamma(\gamma - 1)\Omega\tau_{\text{cool}}}.$$

Note that the above equation is simply a restatement of the energy balance equation, where the viscous heating term on the left-hand side is balanced by the cooling term on the right-hand side. A relation like the one described above necessarily holds whenever self-gravity provides the necessary heating to reach thermal equilibrium, and if transport can be described viscously. The practical inconvenient of Eq. (68) is due to the fact that in general the cooling time in a given configuration is not known *a priori*, and it requires a careful evaluation of the cooling processes in order to determine the amount of stress in a specific case.

We still need to verify whether the condition for the validity of Eq. (68) described above (i.e. that the transport induced by self-gravity is viscous) does hold. It is here that numerical simulations can provide the missing piece of information. Indeed, from the results of the simulation one is able to directly measure the relevant torques and stresses and compare the results with the models described above. The ‘viscosity’ can be estimated from the simulations in two different ways. First, one can compare the secular evolution of the surface density as seen in the simulation with the one predicted for example from Eq. (35) and fit the value of the viscosity to the observed evolution. This has been done [56, 57] for isothermal or polytropic equations of state and the evolution appears to be reproduced reasonably well with values of  $\alpha_{\text{sg}}$  in the range of 0.01-0.03. It should be noted that these simulations had relatively massive discs, with  $M_{\text{disc}} \approx M$ . However, such estimates cannot be directly compared to what should be expected based on Eq. (68), since these simulations, being isothermal, do not have a well defined cooling time. In some cases [57], the simulations require the adoption of a spatially varying  $\alpha_{\text{sg}}$ , prompting the authors to speculate that this is a consequence of a ‘failing’ of the  $\alpha$ -prescription due to the global nature of transport in their models. Now, while global transport might well operate in the massive ( $M_{\text{disc}} \approx M$ ) discs considered in those simulations, we note here that, as already remarked in Section 2.5 above, the simple fact that  $\alpha$  is not constant does not call into question neither the use of the  $\alpha$ -prescription,

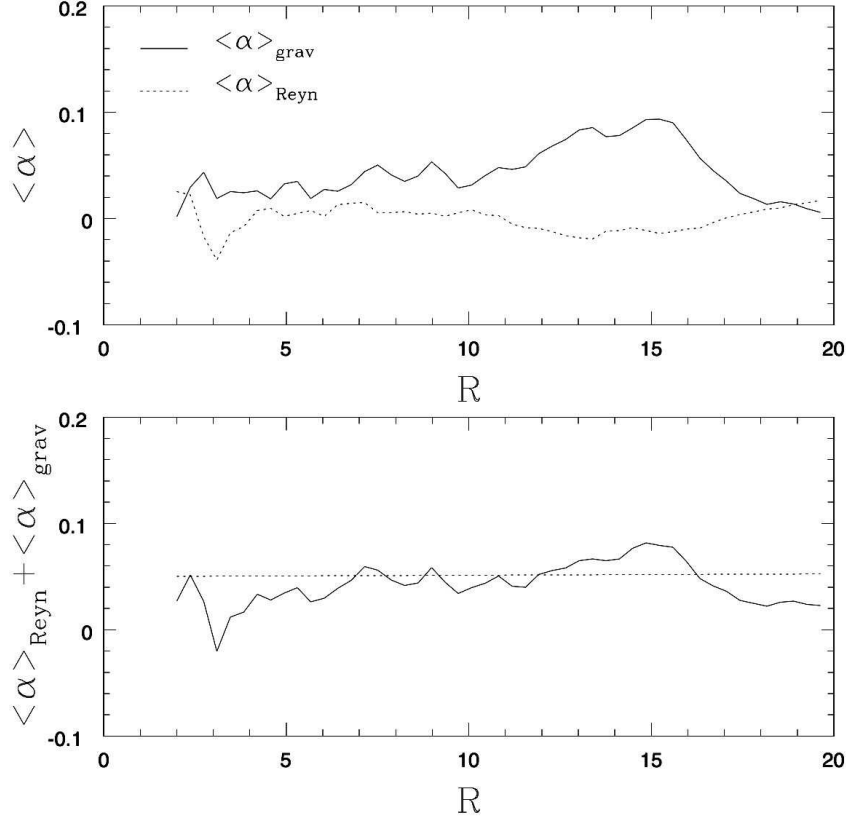


Fig. 9. – Upper panel: gravitational (solid line) and Reynolds (dotted line) contributions to the stress as a function of radius for the simulation shown in Fig. 5. Lower panel: comparison between the total (Reynolds plus gravitational) stress (solid line) and the expected value from Eq. (68).

nor the use of a local viscosity or a diffusion equation to describe the disc evolution.

A second possibility to estimate viscosity is to compute directly the torque exerted by the non axisymmetric disturbances developing in self-gravitating discs. As we will see, this will lead to more insight on the issue of whether transport can be described in terms of local quantities or rather global effects invalidate such a treatment. The starting point for this analysis is Euler equation for a dissipationless fluid, Eq. (3). Consider the case where the fluid velocity is made up of two contributions, one from a (secularly varying) average component, that we call  $\mathbf{v}$ , and one from a rapidly varying fluctuating component, that we call  $\mathbf{u}$ . In a similar way, the self-gravitating potential can be divided in a quasi-steady component and a fluctuating one. It can be shown (see, for example, [31, 78]) that in cylindrical coordinates, the vertically integrated Euler equation takes the following form:

$$(69) \quad \frac{\partial}{\partial t} (\Sigma R v_\phi) + \frac{1}{R} \frac{\partial}{\partial R} (R v_R \Sigma R v_\phi) = - \sum_i \frac{1}{R} \frac{\partial}{\partial R} \left( R^2 \Sigma \langle u_R^{(i)} u_\phi^{(i)} \rangle \right).$$

In Eq. (69) the angle brackets indicate a vertical and azimuthal average and the summation extends to the various fluctuating fields that contribute to the stress. In the case considered here, the two relevant fields are the velocity field and the gravitational field. The stress associated with the fluctuating velocity is usually called the “Reynolds” stress and is given by

$$(70) \quad T_{R\phi}^{\text{Re}} = -\Sigma \langle u_R u_\phi \rangle,$$

while the term associated with self-gravity is given by

$$(71) \quad T_{R\phi}^{\text{g}} = -\Sigma \langle u_R^{\text{g}} u_\phi^{\text{g}} \rangle,$$

where  $\mathbf{u}^{\text{g}} = \mathbf{g}/\sqrt{4\pi G\rho}$ , and  $\mathbf{g} = -\nabla\Phi$  is the gravitational field [77]. Equation (69) shows that in the presence of a non-zero correlation between the radial and azimuthal components of any of the fluctuating fields described above, angular momentum can be removed from the mean flow. Indeed, comparing Eq. (69) to Eq. (33), it is readily seen that such correlations play exactly the same role as a viscous stress tensor for what concerns angular momentum conservation. Of course, we can also use the  $\alpha$ -prescription, that is, we can measure the stress in units of the local vertical integrated pressure:

$$(72) \quad \alpha_{\text{sg}} = \left| \frac{d \ln \Omega}{d \ln R} \right|^{-1} \frac{\langle u_R u_\phi \rangle + \langle u_R^{\text{g}} u_\phi^{\text{g}} \rangle}{c_s^2}.$$

The above formula is simply a measure of the torque provided by non-axisymmetric structures in units of the local pressure and obviously it can always be used to characterize the angular momentum transport due to self-gravity. The question of whether this can be interpreted as a ‘local’ viscosity or not requires a consideration of two further issues: (a) Is the value of  $\alpha_{\text{sg}}$  determined locally or does it depend on the global disc configuration? (b) Does this process dissipate energy as a viscous term would do, that is according to Eq. (50)?

Both questions above can be answered using the results of numerical simulations. Indeed, we can simply measure, based on the simulations,  $\alpha_{\text{sg}}$  from Eq. (72). As we have seen above, in many cases the disc reaches a self-regulated state in which thermal equilibrium holds. We can then compare the torque computed from Eq. (72) with the expectations based on Eq. (68), that is derived under the assumption that the cooling term is balanced by local viscous heating. This exercise has been done in a number of papers [45, 51, 61] and the general results is that the two expressions agree relatively well over a large radial range. Note that the simulations reported in [61] were set up to be local (i.e. they only described a small patch in the disc) and there is therefore no surprise that possible global effects did not show up. On the other hand, the simulations of [45] and those of [51] are global models of the disc (which in these two cases is taken to be of relatively low mass  $M_{\text{disc}}/M \ll 1$ ) and global effects could in principle be present. Fig. 9 shows an example of such calculation, where the stress computed from Eq. (72), (solid line in the lower panel) is compared to the expectation from Eq. (68) (dotted line) for the simulation shown in Fig. 5. An even clearer example is shown in Fig. 13 of [51], where the same comparison is made, again showing a remarkable agreement over a large portion of the disc. Note that the simulations of [45] and those of [51] employ significantly different cooling prescriptions, in that while in [45]  $\tau_{\text{cool}}\Omega$  is set to be a constant, resulting in an approximately constant value of  $\alpha_{\text{sg}}$  (cf. Eq. (68)), in [51]  $\tau_{\text{cool}}$  is either constant



or self-consistently computed based on the solution of the radiative transfer equation, therefore resulting in a non constant value of  $\alpha_{\text{sg}}$ . In all cases, however, energy appears to be dissipated according to the expectations from a viscous model. Clearly, such a comparison is only meaningful when the disc is in thermal equilibrium. As discussed above, for very massive discs [46], where  $M_{\text{disc}} \lesssim M$ , a quasi-steady thermal equilibrium is generally not reached and this kind of comparison cannot be made.

It is also possible to estimate whether the value of  $\alpha_{\text{sg}}$  is determined locally or not. For example one can calculate the contribution to the gravitational stress at a given reference location coming from a region of finite size around it [45]. It can then be seen that the dominant contribution to the stress arise from a neighbouring region of size  $\approx 10H$ . A similar result has also been obtained using a different method for local simulations [61]. The local approximation thus breaks down if  $10H \gtrsim R$ , or  $H/R \approx M_{\text{disc}}/M \gtrsim 0.1$ . Note once again the different behaviour of light and massive discs in this respect.

We have thus seen that, at least in the case of low mass discs, the transport of energy and angular momentum induced by gravitational instabilities can be well reproduced in terms of a local, ‘viscous’ model. Is this result to be expected? Low mass discs can support a relatively large spectrum of density waves, with different  $m$  and with wavenumbers of the order of  $1/H$ . In the presence of effective cooling, many of such waves can be easily excited. One possibility is that such waves interact with one another, undergo shocks and dissipation, and therefore do not propagate over large radial distances, especially if their typical wavelength is small. In such a case, a local description of the transport induced by the waves would appear to be appropriate. Alternatively, another possible outcome is that a relatively small number of waves can propagate significantly and establish a pattern of global modes, which persist over a few orbits and extend over a large radial range. Such persistent, global modes have been observed in some hydrodynamical simulations of accretion discs, even in the case of low mass discs [51], but even in this case the calculated stress tensor agrees with the value required by energy balance in a viscous disc.

The energy flux transported by self-gravitating density waves can also be calculated analytically [76]. It can be shown [79] that in general the energy flux (contrary to what happens for a viscous process) is not proportional to the stress tensor, and that some other, ‘anomalous’ terms are present, which would preclude a local treatment. These terms can be evaluated within a WKB analysis and it is shown [79] that they vanish at corotation (where the pattern speed of the wave  $\Omega_{\text{p}} = \Omega(R)$ ), their contribution increasing as we move away from corotation and becoming significant if  $|\Omega_{\text{p}} - \Omega(R)| \approx \Omega(R)$ . Let us consider the case of an approximately Keplerian disc. In this case, for a wave generated close to corotation such global terms become important only if the wave has travelled a radial distance of the order of  $\Delta R/R \approx 2^{2/3} \approx 1.58$ . Now, it is plausible to assume that the minimum distance travelled by a density wave is no less than its radial wavelength  $\approx 2\pi H$ . A necessary (but not sufficient!) condition for global effects to be negligible is thus  $H/R \lesssim \Delta R/2\pi R \approx 0.25$ . Once again, we see that thin and thick discs (or, equivalently, light and massive discs) behave in a significantly different way. While for thick and massive discs a local approach appears inadequate, in the case of lighter and thinner discs it is possible to have either a local or a global behaviour, depending on the coherence of the spiral modes involved. The numerical evidences described above appear to imply that a local approach is generally valid for thin discs, and indeed even in the cases where a global pattern is observed, its radial extent is not large enough to prevent the use of a local model [51].

Let us summarize what we have discussed so far. The behaviour of self-gravitating

discs depends on three dimensionless parameters: the parameter  $Q$  (which measures how hot the disc is), the parameter  $M_{\text{disc}}/M$  (which measures how massive the disc is) and the parameter  $\Omega\tau_{\text{cool}}$  (which measures how fast does the disc cool). If  $Q$  is significantly larger than unity the disc is gravitationally stable, while if  $Q$  approaches unity it becomes unstable. For an approximately Keplerian rotation the condition  $Q \approx 1$  is equivalent to requiring  $M_{\text{disc}}/M \approx H/R$ . The behaviour of a gravitationally unstable disc depends on  $M_{\text{disc}}/M$  and  $\Omega\tau_{\text{cool}}$ . In particular, if  $\Omega\tau_{\text{cool}} \lesssim 1$  the disc rapidly undergoes fragmentation and produces a number of gravitationally bound clumps, while for larger values of the cooling rate fragmentation is inhibited. In the latter case, if  $M_{\text{disc}}/M \lesssim 0.25$ , the disc settles down in a quasi-steady self-regulated state, the instability takes the form of a tightly wound spiral structure and the transport of energy and angular momentum induced by it are usually well described within a local ‘ $\alpha$ -like’ approach. On the other hand more massive discs, with  $M_{\text{disc}}/M \gtrsim 0.25$ , cannot be described within the local approximation and global transport is expected to occur. In this case, a quasi-steady self-regulated state is not generally observed in the simulations (see Fig. 7), which show instead a series of recurrent strong episodes of instability, during which the gravitational torques can be very large ( $\alpha_{\text{sg}} \approx 0.5$ ), and determine a fast redistribution of the gas within the disc.

## 6. – THE RELATION BETWEEN FRAGMENTATION AND TRANSPORT

The above discussion on the transport induced by self-gravity in a light disc (i.e. for the case where a local description in terms of an  $\alpha$  viscosity is valid) allows us to give a more detailed look at the issue of fragmentation. Let us then concentrate on light, quasi Keplerian discs. If these discs are self-gravitating, they settle down in a quasi-steady self-regulated state, with  $Q \approx \bar{Q}$ , in which they transport angular momentum with a torque strength that can be expressed in terms of  $\alpha$  through Eq. (68). However, from the discussion of Section 4 we know that if the cooling rate is smaller than a given threshold  $\tau_{\text{cool}}\Omega < \beta_{\text{crit}} \approx 3 - 6$ , the disc undergoes fragmentation. Therefore, through Eq. (68), a minimum value of  $\Omega\tau_{\text{cool}}$  to prevent fragmentation corresponds to a maximum value of  $\alpha_{\text{sg}}$ . It is then possible to interpret the fragmentation threshold in a different way, in terms of how much angular momentum can be transported through self-gravity in a steady state. This interpretation is further reinforced by numerical simulations which employ different values of the ratio of specific heats  $\gamma$  [65]. It appears that as  $\gamma$  decreases, discs are more susceptible to fragmentation. For example, while for  $\gamma = 5/3$ , the threshold for fragmentation is at  $\beta_{\text{crit}} \approx 6.5$ , it increases to  $\beta_{\text{crit}} \approx 12.5$  for  $\gamma = 7/5$ . The results of this analysis is shown in Fig. 10, where the three curves show the relation between gravitationally induced viscosity  $\alpha$  and the cooling time  $\tau_{\text{cool}}$  in thermal equilibrium, expressed in Eq. (68), for three different values of  $\gamma = 2$  (solid line),  $\gamma = 5/3$  (short dashed line) and  $\gamma = 7/5$  (long-dashed line). The green squares indicate the value of  $\tau_{\text{cool}}$  at which fragmentation is observed in [65], while the triangle refers to the 2D simulations with  $\gamma = 2$  presented in [61]. It is then interesting to note that fragmentation occurs whenever the value of  $\alpha$  needed to be provided by self-gravity in order to reach thermal balance exceeds a maximum value of order  $\alpha_{\text{max}} \approx 0.06$ . For relatively long cooling times, the strength of the torques induced by self-gravity does not need to be very large in order to provide the required heating to reach equilibrium. For shorter cooling times, the saturation amplitude of the spiral structure increases until it becomes too large to be sustained by the disc, which then undergoes fragmentation. The torque provided by the spiral structure at this maximum sustainable amplitude corresponds to

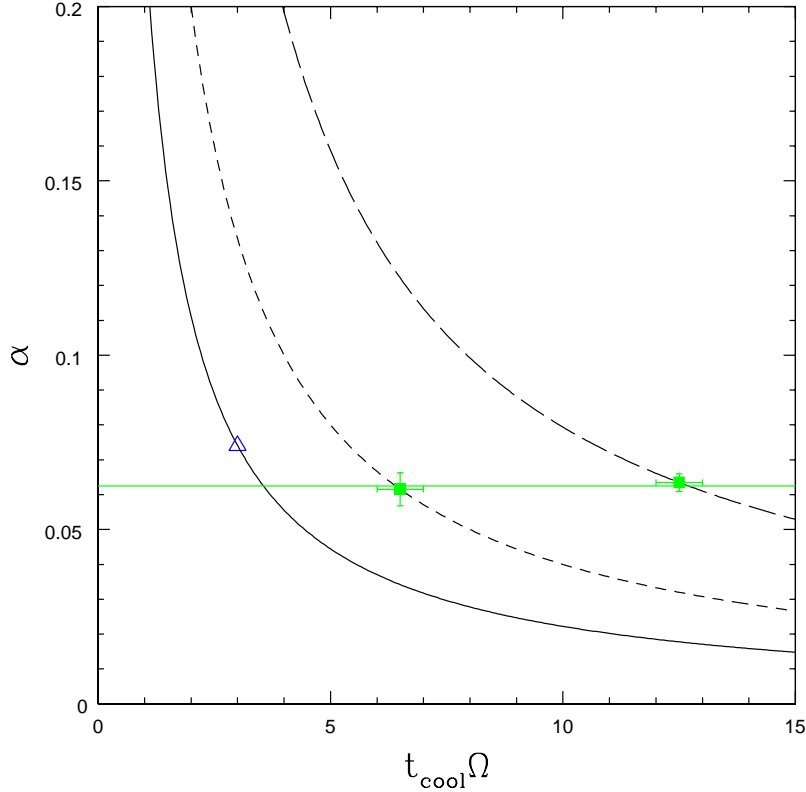


Fig. 10. – The relation between gravitationally induced viscosity  $\alpha$  and the cooling time  $\tau_{\text{cool}}$  in thermal equilibrium, expressed in Eq. (68), for three different values of  $\gamma = 2$  (solid line),  $\gamma = 5/3$  (short dashed line) and  $\gamma = 7/5$  (long-dashed line) (from [65]). The green squares indicate the value of  $\tau_{\text{cool}}$  at which fragmentation is observed in [65], while the triangle refers to the 2D simulations presented in [61]. It appears that fragmentation occurs whenever the value of  $\alpha$  needed to be provided by self-gravity in order to reach thermal balance exceeds a maximum value of order  $\alpha_{\text{max}} \approx 0.06$ .

$\alpha_{\text{max}} \approx 0.06$ .

It is important to stress that the threshold value for  $\alpha$  discussed above refers to a quasi-steady state, in which the disc stays in thermal equilibrium and the relevant physical quantities do not change significantly on time scales shorter than the thermal timescale. This does not happen for massive discs (with masses comparable to that of the central object), that can generate transient strong spiral episodes, with correspondingly large values of the stress  $\alpha$ , which, however, do not last for longer than one dynamical timescale.

If the fragmentation threshold is ultimately due to an inability of the disc to redistribute angular momentum on short timescales for a prolonged period of time, then it is possible to envisage a further possible route to fragmentation, in cases where the disc, even if the cooling timescale is long, is fed at a high accretion rate. This possibility, although it may play an important role in several contexts, has not yet been tested nu-

merically. In particular, the maximum accretion rate sustainable at any given radius in the disc is only dependent on the local temperature and is given by  $\dot{M}_{\max} = 2\alpha_{\max}c_s^3/G$ .

## 7. – APPLICATION TO OBSERVED SYSTEMS: STAR AND PLANET FORMATION

The discussion of the sections above has focussed on the basic physical processes that determine the evolution of self-gravitating accretion discs, without referring to any specific astrophysical system where such mechanisms can be at work. We have thus seen what are the basic parameters involved in the determination of the disc properties and evolution. We now turn to an examination of a few concrete cases and we will try to connect the theoretical picture described above to the observed properties of these systems.

**7.1. Protostellar discs.** – The formation of protostellar discs is the natural outcome of the process of star formation. Stars are supposed to form from dense condensations within turbulent molecular clouds, called molecular cloud cores (see [80] for a recent review). The typical specific angular momentum of one such core is of the order of  $10^{21}\text{cm}^2/\text{sec}$  and is therefore much larger than the specific angular momentum of a star (for comparison, the Sun has a specific angular momentum of the order of  $10^{16}\text{cm}^2/\text{sec}$ ). The gas that is going to form a star therefore has to get rid of most of its angular momentum. The collapse of a molecular cloud core is thus bound to form initially a quite extended flattened object, a disc, with sizes of the order of tens to hundreds of AU. Angular momentum redistribution inside the disc allows accretion to proceed and to feed the forming protostar. Clearly, during this process the mass balance in the star-disc system moves from a disc-dominated initial stage, to a star-dominated evolved stage, after the main infall phase has finished. The effects of the disc self-gravity are then going to be more important for younger object than for evolved ones. This evolution has an observational counterpart in a frequently used classification [81], which identifies as Class 0 and Class I the youngest objects, which are still embedded into a dusty gaseous envelope. During this phase the disc is still actively accreting from the envelope and the disc mass is expected to be very large. Unfortunately, during this phase most of the activity which occurs in the disc is enshrouded by the envelope and is difficult to observe, with the exception of a few cases. Class II objects have cleared out most of the envelope and the protostar-disc system is now visible. Such objects are often called Classical T Tauri stars. Since the main infall phase has ended, disc masses in the T Tauri phase are expected to be smaller than in the Class 0 and Class I case. Finally, Class III objects are the most evolved class of protostars, where the gaseous disc has almost entirely dispersed, leaving a remnant ‘debris disc’ mostly formed of rocks and pebbles (akin to the Zodiacal dust in our Solar System).

The properties of protostellar discs are obviously a strong function of the stellar mass. Most of the data available at the moment refers to discs around solar type stars. However, in recent years a significant amount of data has started to become available for discs at the extremes of the stellar mass distribution, that is for brown dwarfs at the low mass end, and for massive O and B type stars at the upper end of the spectrum. In the latter case, in particular, there is strong evidence pointing to the presence of very massive discs, and these objects are therefore an interesting laboratory to test the models described in the previous sections. We will therefore discuss both the well known case of solar type protostars and the case of massive O-B protostars.

**7.2. Solar mass young stars.** – Disc masses in the T Tauri phase are generally estimated from sub-mm emission, since at these wavelengths dust emission is usually optically thin and we can thus easily convert the observed flux into a mass, if we know what is the dust opacity [16]. Recent surveys of protostellar discs around solar mass stars in the Taurus-Auriga complex [82] indicate disc masses between  $10^{-4}M_{\odot}$  and  $10^{-1}M_{\odot}$ , with a median value of  $5 \cdot 10^{-3}M_{\odot}$ . In another case, [83] the results for a smaller sample (possibly biased to larger masses) indicate a median mass one order of magnitude larger and in a few cases discs as massive as  $0.2M_{\odot}$  have been found. Similar values are also reported for the Orion region, where in a few cases masses as large as  $0.4M_{\odot}$  have also been reported [84]. Since the stellar mass for these samples ranges from 0.1 to  $1M_{\odot}$ , the mass ratios between the disc and the star is between  $10^{-3}$  and a few times  $10^{-1}$ . The thickness of protostellar discs can be estimated from measurements of the temperature in the disc, and it turns out that in general  $H/R \approx 0.1$ . Bearing in mind that gravitational instabilities develop when  $M_{\text{disc}}/M_{\star} \sim H/R$ , we see that objects in the upper end of the mass distribution for T Tauri stars are expected to be gravitationally unstable.

It should be noted, however, that such measurements suffer from very large systematic errors, mostly due to uncertainties in the dust opacity. Indeed, in many cases there is evidence for relatively evolved dust grains, so that the dust size distribution can be significantly different than in the interstellar medium [85,86]. If dust grains have evolved to relatively large sizes, their opacity would be reduced and the inferred disc mass would correspondingly increase. Additionally, it is important to stress that such measurements are sensitive only to the dust content of the disc, which is expected to be a very small fraction of the total gas mass (in the interstellar medium, for example, the dust to gas ratio is of the order of 0.01). The dust mass measurements are then rescaled to obtain the gas mass, by adopting the standard interstellar medium scaling factor, the applicability of which is highly uncertain. In summary, it is likely that most of the above estimates are actually underestimating the real disc masses [32]. We thus see that even in many T Tauri objects gravitational instabilities might contribute substantially to the angular momentum transport. The importance of self-gravity becomes even higher when one consider that T Tauri stars are expected to have already accreted much of their mass from the disc. If by this rather evolved stage self-gravity is still important, it must have dominated the dynamics of younger systems, such as Class I objects.

Accretion rates are generally measured from the strength of emission lines emitted as the gas reaches the star in the so called ‘magneto-spheric accretion’, when the innermost parts of the disc are truncated by the stellar magnetic field, and the gas falls almost freely onto the star along magnetic field lines, or from the veiling of photospheric lines due to accretion shocks [87,88]. It should be noted here that these measurements refer to the accretion rate onto the star, which does not need to correspond to the accretion rate at larger distance, if significant sinks of matter are present, such as a disc wind (which is generally very small) or, more likely, as a massive planet, whose presence might act like a dam for the accretion flow [89]. In any case, typical values range between  $10^{-9} - 10^{-7}M_{\odot}/\text{yr}$ . With such relatively low values of the accretion rate, the expected disc luminosity is relatively small and the energy balance in the disc is determined mostly by irradiation from the central star [39]. While the detailed balance of course might vary from object to object, it is generally hard to look for accretion signatures in the broadband SED of T Tauri stars.

What about the role of self-gravity? Theoretical models of the evolution of protostellar discs clearly indicate that in the early phases of star formation (in the Class I phase) the disc is massive enough to be self-gravitating and the transport should be dominated by

self-gravity [32-34]. Such effects are expected to become less important in the subsequent T Tauri phase.

From the observational point of view, an intriguing possibility would be to directly observe a spiral morphology from high angular resolution observations. In the near future, the Atacama Large Millimeter Array (ALMA), might in principle detect such detailed structures in the dust continuum. We have seen in section 4 above that for low mass discs (such as those that might be expected around T Tauri stars), the spiral structure is tightly wound and characterized by high- $m$  and short wavelengths (see Fig. 6). This can be difficult to observe even at the high resolution achievable with ALMA. On the other hand, more open features, typical of more massive discs, might be detected more easily. In a couple of cases such extended spiral structure has been observed, in the case of AB Aur (an evolved intermediate mass protostar) [90, 91] and in a case of a young Class 0 object [92]. The latter case is particularly interesting, since estimates of the disc mass indicate that the disc makes up more than 30% of the system. On the other hand, in the case of the more evolved AB Aur, the disc mass is very low and estimates of the disc surface density and temperature indicate  $Q \approx 10$ , which is too high to argue in favour of a gravitational instability.

**7.3. *FU Orionis* objects.** – A remarkable group of solar mass young stars are the so-called FU Orionis objects. FU Orionis objects are a small class of protostellar systems undergoing large outbursts, during which their luminosity increases by as much as three orders of magnitude. It is currently believed that the origin of the outburst lies in a sudden enhancement of the accretion rate in the disc, that can rise to values of the order of  $10^{-4} M_{\odot}/\text{yr}$ . Unfortunately only a very small number of such systems have been observed and in only three cases (the prototypical FU Ori, V1057 Cyg and V1515 Cyg) we have a detailed knowledge of the light curve over a long period of time [93], which enables us to have some insight on their evolution and therefore on the dynamics of the outburst.

FU Orionis objects are generally surrounded by dusty envelopes [94], and substantial infall from the envelope, at a rate of approximately  $10^{-6} M_{\odot}/\text{yr}$  is often required by detailed modeling of the outburst [29, 30]. These characteristics tend to identify FU Orionis objects as transient episodes of enhanced activity in an otherwise quiescent Class I object.

The disc origin of the large luminosities observed in these systems has been confirmed recently by high-resolution interferometric observations [95], which indicate that emission comes from an extended (a few AU) source, like a disc. Additionally, it has been shown [96, 97] that the spectral energy distribution of FU Orionis objects is to first approximation consistent with the one expected from a standard steady state accretion disc (as sketched, for example, in Fig. 3). Indeed, this is essentially the only clear case where a protostellar disc displays the typical emission feature of an active disc. Detailed modeling [97] shows that the typical accretion rates are of the order of  $10^{-4} M_{\odot}/\text{yr}$ . The agreement however only holds at relatively short wavelengths, in the optical and near infrared, while in the mid infrared a substantial excess emission with respect to the expected disc spectrum is observed. Given the youth of the system, we expect that the disc mass in these cases should be relatively large and it is therefore plausible to associate the excess emission at mid-infrared wavelengths with the effect of self-gravity. Indeed, self-regulated discs are hotter in their outer parts, due to the combined effect of the self-regulation process and of the flattening of the rotation curve when the disc mass is large. As a result, the SED tends to show some excess emission at long wavelengths (see

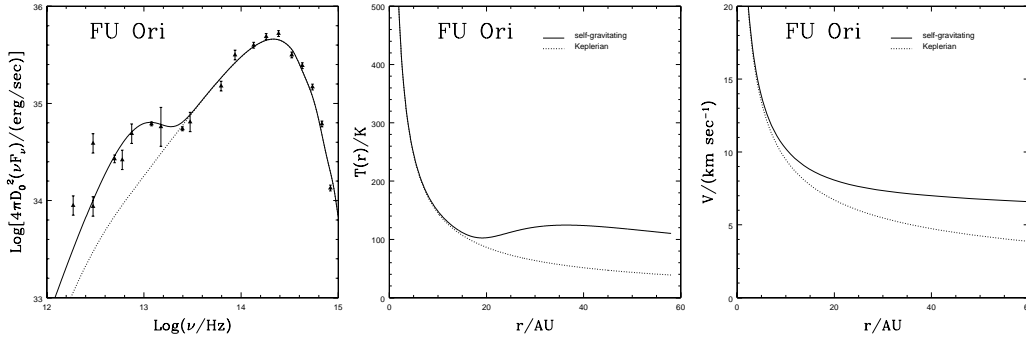


Fig. 11. – Modeling of FU Ori SED with a self-regulated self-gravitating disc. Left panel: observed (data points with error bars) and model SED (solid line). Middle panel: temperature profile implied by the self-regulated model (solid line) compared to a non-self-gravitating model with the same parameters (dashed line). Right panel: Rotation curve of the model (solid line) compared to a Keplerian curve (dashed line).

Fig. 4), consistent with observations of FU Orionis objects. Detailed modeling [49] has shown how it is possible to fit accurately the long-wavelength SED of the best studied FU Orionis objects, based on steady-state models of self-regulated discs (note that while clearly these outbursting objects are evolving, the evolution timescale is longer than the dynamical timescale, so that a steady-state model is reasonably appropriate). An example of such modeling is shown in Fig. 11 and refers to the prototypical case of FU Ori (full details can be found in [49]). These models do however require fairly massive discs, with  $M_{\text{disc}} \approx M_*$ . Such models predict deviations from Keplerian rotation, that might be detected from observations of the line profiles emitted by these systems [98]

Traditional models to explain the cause of the outburst assume that this is due to a thermal instability in the disc, caused by the steep change in disc opacity as hydrogen becomes ionized [29, 99]. Such models have been extensively tested against observations and reproduce the broad features of the light curve. However, agreement is only obtained for rather extreme values of the parameters involved (for example, the  $\alpha$  viscosity parameter is required to be as low as  $10^{-4} - 10^{-3}$ ) and by introducing some *ad hoc* triggering event [30, 100].

An alternative outburst model is that the increase of mass accretion rate is due to the development of a gravitational instability. As we have seen above, SED modeling is consistent with the presence of fairly massive discs. Now, when the disc mass is of the order of the central object mass, we are in the regime where we expect the development of large scale global gravitational instabilities (cf. Section 5). Numerical simulations of this regime [46, 101] show that accretion can be highly variable, and indeed characterized by periods of quiescence followed by recurrent episodes of enhanced accretion. A gravitational origin for FU Orionis outburst has been proposed [102] (or a combination of gravitational and MHD instabilities, [103]), even though a detailed comparison of the model light curves with observations is still lacking.

**7.4. Massive protostars.** – In the previous two subsections, we have concentrated on the properties of relatively low mass young stars, i.e. stars with a mass of the order of one solar mass. Stars of higher mass behave in a significantly different way. One of the

main reasons behind this different behaviour lies in the fact that the typical contraction time of the forming protostar (that is roughly the time it takes for a forming star to ignite nuclear reactions in its center) is much shorter for high mass stars than for low mass stars. In particular, for stars of mass larger than  $\approx 8M_{\odot}$ , the star reaches the main sequence while still accreting matter from the surrounding protostellar core [104]. Hence, while for low mass stars we can observe optically visible pre-main-sequence stars as T Tauri (Class II) stars after the main infall phase has terminated, which are therefore surrounded by a relatively low mass disc, this is not the case for stars of mass larger than  $8M_{\odot}$ . In the latter case, the protostar is always going to be found during the main infall phase, and we thus expect that, if discs are present also here, they should be characterized by much larger masses than in T Tauri stars, and their properties should reflect more those of younger protostars (Class 0 or Class I).

Clearly, observations of high mass protostars and of their circumstellar environment (including the possible presence of discs) are difficult, due to the fact that they are still embedded within their parent core. Nevertheless, observations in the last few years have been able to reveal a significant number of them, and there is now a substantial evidence that protostellar discs are present also for high mass stars (see [105] for a recent review). It should be mentioned however that the evidence for the presence of discs is mostly circumstantial. In most cases, the ‘disc’ is found as a structure that lies perpendicular to a large scale outflow, which is common around young stars, and is easy to detect. Such structures are either seen in infrared continuum observations as dark silhouette discs against a bright background [106], or through observations of thermal [17, 18] and non-thermal (usually maser) emission lines [107–110], which have the advantage of providing information about the velocity structure of the circumstellar environment, so that a disc can be associated with a velocity gradient perpendicular to the outflow axis, usually interpreted as a sign of rotation.

The general result is consistent with the expectations outlined above. The discs are indeed found to be very massive, with masses comparable to, and in some cases even larger than, that of their central star. Although the data are still very uncertain, there is also some evidence for non-Keplerian rotation, as expected when the disc mass is high [18]. Observed structures can be broadly divided in two groups: some are very extended (with sizes of  $\approx 10^4$  AU), very massive (with masses much larger than their central stars) and the luminosity of the central object is generally consistent with an O-type protostar (mass larger than  $20M_{\odot}$ ). The second group of objects is characterized by smaller sizes ( $\sim 10^3$  AU), smaller mass ( $M_{\text{disc}} \lesssim M$ ), and are generally associated with B-type stars (with masses between 8 and  $20 M_{\odot}$ ). Objects in the first group are so large that their typical orbital timescale is of the order of the lifetime of the structure. Such objects are therefore unlikely to be in centrifugal equilibrium and is therefore difficult to identify them as accretion discs (even though a smaller size disc can in principle be hidden within these structure). On the other hand, objects in the second group may be regarded as proper discs.

Let us now consider what accretion rates are associated with the discs in the second group of objects. Here again, observations are very difficult. A well constrained parameter is the outflow rate, which is produced in the inner parts of the disc (at a distance  $\lesssim 1$  AU) from the star, and it is of the order of  $10^{-4} - 10^{-3} M_{\odot}/\text{yr}$ . Clearly, the accretion rate through the disc should be at least of this order of magnitude to feed the outflow. Can the disc sustain such large accretion rates? Let us first consider the case where transport can be simply described in terms of a local, viscous phenomenon. As we have seen above, this is the expectation when the transport is induced by a gravitational insta-



bility in low mass discs. We have also seen that in this case a quasi-steady, self-regulated spiral structure does not provide a torque larger than  $\alpha \approx 0.06$  (when measured in terms of the  $\alpha$ -prescription) without undergoing fragmentation. It is then easy to estimate the accretion rate expected under these conditions, from  $\dot{M} \approx 2\alpha c_s^3/G$ , where  $c_s$  is the sound speed. Typical temperatures in the outer disc for these objects are of the order of 170K (this is the case of IRAS 20126+4104, the best studied among these objects [18]), which correspond to a maximum accretion rate of  $\dot{M} \approx 2 \cdot 10^{-5} M_\odot/\text{yr}$ , well below the rate needed to steadily feed the outflow from the inner disc. On the other hand, we have to keep in mind that the picture where gravitational instabilities provide a local, viscous source of angular momentum transport is only appropriate for relatively low mass discs. Here, the disc masses are much higher and hence the situation is more similar to the case shown in Fig. 7, than to the one shown in Fig. 6. We should thus expect the development of a series of recurrent global spiral structures, which can temporarily (for a period of the order of the outer dynamical timescale, which in this case is of the order of  $10^4$  years) transport angular momentum much more efficiently than a local process would do. As discussed above, during such episodes the stress tensor induced by the spiral structure can be up to a factor ten larger than in a quasi-steady self-regulated case, and we can thus reconcile in this way the observationally implied accretion rates with the rates expected from theoretical arguments.

**7.5. Planet formation.** – The process of planet formation is intimately linked to the process of star formation. The interest in this subject has grown significantly in the last ten years, and in particular after the discovery of the first planetary system around a solar type star [111], outside our own Solar System. To date, hundreds of extra-solar planets have been discovered, and the list is ever-growing. Dedicated space missions, such as Darwin and the Terrestrial Planet Finder, are going to be launched in the near future. This large set of data allow us to compare theoretical models of planet formation with a statistical sample, rather than to the only one example (the Solar system) that was available until very recently. Our theories about planet formation are going to be significantly improved and modified as new data become available.

The properties of extra-solar planets appear to be very different from the ones of the planets in our Solar System, although in many cases they reflect the limitations intrinsic to the detection methods used. In fact, most extra-solar planets are discovered through the ‘radial velocity’ technique, based on the measurement of the radial velocity induced by the presence of the planet on the motion of the host star. This method naturally favours the detection of planets of high mass and large eccentricities, which would induce the largest radial velocity. Also, planets at large separations, which have very long periods, cannot be detected this way. An additional observational limitation is given by the fact that the radial velocity provides information about  $M_p \sin i$ , where  $M_p$  is the planetary mass and  $i$  is the inclination of the planetary orbit. If (as it often happens) the system is not eclipsing, we do not have information on the inclination and thus the planetary mass is only known to within this potentially large unknown factor. In any case, the inferred planetary masses range from 5 Earth masses, in the case of the ‘super-Earths’ recently discovered [112], up to 20 Jupiter masses, much larger than the most massive planets in the Solar System. It should be noted that many of these massive planets are found at very small separations from their host stars, smaller than 0.1 AU, where the locally available mass is too small to form the planet, thus suggesting that during the course of the planet formation substantial radial migration can happen. Additionally, while in the Solar System the orbital eccentricity is generally low, below

0.1 (except for Mercury and Pluto), extra-solar planets essentially span all the possible eccentricities, reaching values as high as 0.9.

Parallel to the investigation of extra-solar planets, we have also improved our knowledge of the structure of planets within the Solar system. Particularly relevant to the discussion on theories of planet formation are the most recent evidence for the presence of a solid core in the interior of Jupiter and Saturn [113]. In fact, the estimates of the mass of this solid core have been revised and they now appear to be smaller than previously thought. In particular, while for Saturn the core mass is of the order of  $20M_{\oplus}$ , in the case of Jupiter the upper limit is  $14M_{\oplus}$  and models with no core at all are also not excluded.

Currently, there are two models for the formation of planets. The first (and the more widely accepted) is the so called ‘core accretion’ model, while the second (which was usually disregarded, but has recently gained progressively more attention) is the so called ‘gravitational instability’ model. Clearly, in the context of the present paper, the second model appears to be more interesting. However, as we shall see, gravitational instabilities may play an important role even in the core accretion model. While a thorough review of planet formation theories is clearly beyond the scope of this paper, I here briefly summarize the main features of the two models.

**7.5.1. The core accretion model.** In the core accretion model [114] (see also [115] and [116] for two recent reviews), planets are supposed to form sequentially out of the solid component present in protostellar discs. The first step in planet formation is the growth of interstellar dust through inelastic collisions up to a size of tens of centimeters. This growth is seen in observations of protostellar discs, whose opacities often indicates the presence of large dust grains [85], and where high-resolution spectroscopy in the infrared indicate the presence of structured silicate grains [117, 118] (incidentally, the large uncertainties in protostellar disc masses mentioned in section 7.1s are essentially due to the unknown level of dust growth). For sizes larger than one meter, the collisional agglomeration of dust is not effective. The growth in this mass range, up to the formation of kilometer-sized planetesimals, is still largely unknown (but see below). Once km-sized planetesimals have formed they essentially decouple from the gas and keep growing through their mutual interaction, enhanced by gravitational focussing. In this way a few planetary cores emerge from the background, and if the core reaches a size of the order of  $5 - 10M_{\oplus}$  it is able to further accrete a large gaseous envelope, hence becoming a giant planet, like Jupiter. An important attractive of this model is that it is able to produce both small, Earth-like, rocky planets, and the large Jupiter-like gaseous planets. Note that within this model, Jupiter is bound to have a massive rocky core, the evidence in favour of which, as mentioned above, is not robust. Apart from this, two other problems affect this model for planet formation. First of all, it is a relatively slow process. Detailed models [114] require 10 million years in order to form Jupiter, which is just within the estimates of the lifetime of protostellar discs. A second important problem is related to the formation of planetesimals. It is in this respect that the disc self-gravity might play a fundamental role in planet formation and I thus discuss it here in some detail.

The issue arises when one considers the motion of meter sized solid particles within a gaseous disc. Even in the case where the disc contribution to the radial gravitational field is negligible, the gas rotation can be non-Keplerian, due to the effects of pressure gradients. As discussed in section 2.3, pressure gradients provide additional support to the gas against gravity, so that the rotation curve can be non-Keplerian to some extent (Eqs. (15) and (16)). The degree of non-Keplerianity in this case is determined by

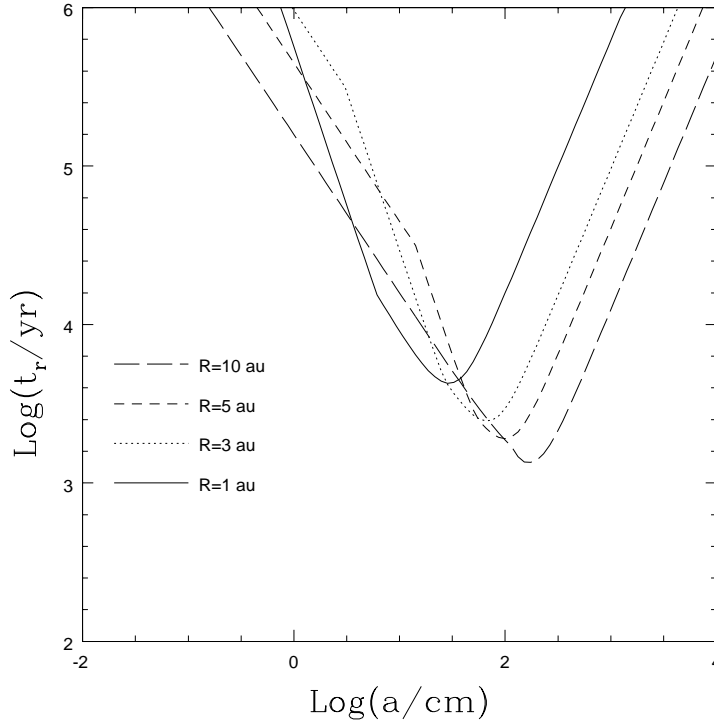


Fig. 12. – Solid particles migration timescale as a function of particle size for a typical protostellar disc configuration. The disc is here assumed to have a smooth pressure distribution, uniformly decreasing outward as a power law. The four lines refer to four different radial locations in the disc. From [21].

$(c_s/v_\phi)^2 \approx (H/R)^2$ . The sign of this contribution, depends on the pressure gradient, so that if the pressure decreases with increasing radius, the rotation is sub-Keplerian. Solid particles, on the other hand, do not feel the effect of pressure and therefore their rotation curve to first approximation is Keplerian. This thus determines a velocity difference between the solids and the gas (note that this effect is analogous to the so called ‘asymmetric drift’ which occurs in galaxy dynamics, cf. [119]). If the solids and the gas are coupled via a drag force (for example, the usual Stokes or Epstein drag, which are usually dependent on the solid particle size) this may have some sizable effect on the solid’s motion. Let us consider a smooth disc with pressure decreasing outwards [20], in this case the solids would move faster than the gas and the drag force therefore acts in such a way to remove angular momentum from them and thus determines an inward radial migration. The magnitude of such radial motion and the timescale for radial migration depends on the particle size. In particular, for very small sizes the solids are so strongly coupled to the gas that they essentially follow the gas motion and their velocity difference with respect to the gas is negligible. For very large sizes, on the other hand, the coupling becomes very weak, and the solids move in essentially Keplerian orbits, with a very small inward velocity. The most interesting behaviour occurs for intermediate sizes, for which the solids are coupled to the gas, but not so strongly to completely follow the

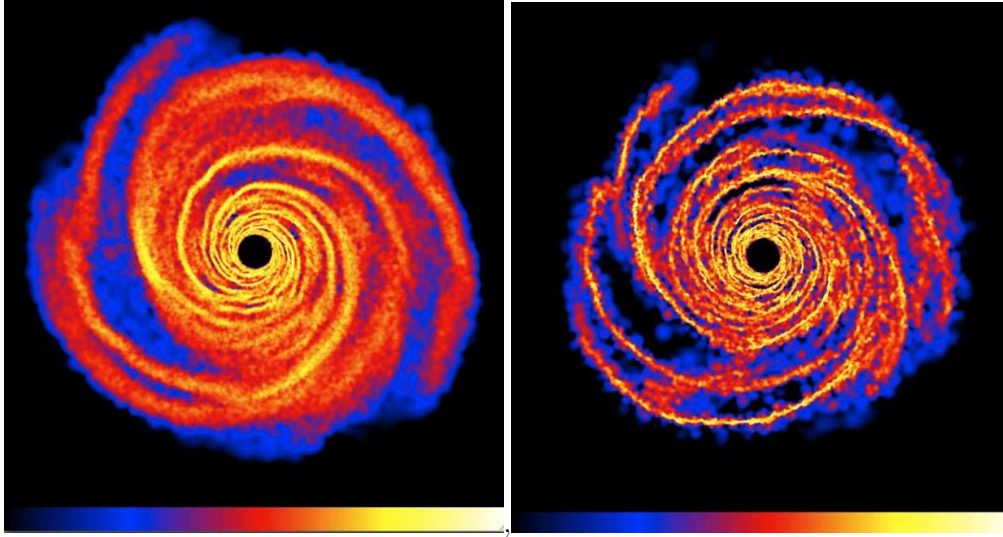


Fig. 13. – Surface density of the solid component in a self-gravitating gaseous discs, where the gas and the solids are coupled through drag. Left panel: solid size equal to 10 meters. Right panel: solid size equal to 50 cm. While in the 10 meter case the solid density is very similar to the corresponding gas density, in the 50 cm case the solid concentrate very efficiently into filamentary structures at the peaks of the spiral arms. From [21].

gas streamlines. The migration timescale as a function of particle size is shown in Fig. 12 for a typical smooth protostellar disc configuration at different radial distances from the star. It can be seen that for sufficiently small and for sufficiently large particle sizes the migration timescale is very long. For meter-sized objects however, the effect becomes stronger and the migration time can be as small as a few thousand years, a very short timescale with respect to the typical solid growth time at these size. This then poses a severe problem for the core accretion model, in that as soon as the dust grows in size and reaches the meter-size region, it suffers from this very fast radial migration and would essentially disappear into the innermost regions of the disc before being able to grow further and form planetesimals.

The above behaviour occurs for a smooth disc, in which the pressure is a monotonic function of radius. If the disc is characterized by a long-lasting spiral structure, in which the density (and hence the pressure) presents local maxima, the situation can be significantly different [120, 121]. Indeed, if the pressure increases with radius, the gas motion is locally super-Keplerian and the direction of the drag force is reverted, leading to a fast outward migration. In turn, in the presence of a local pressure maximum, the solid tend to migrate and concentrate to the maximum. The timescale for this process can be even faster than the migration timescale. First of all, recall that the typical size of a spiral perturbation due to the disc self-gravity is of the order of  $H$ . This means that the pressure gradient is enhanced with respect to a smooth disc by a factor  $R/H$ . Secondly, in order to migrate to the peak of the spiral structure, the solids have to move a distance  $H/R$  shorter than in a smooth disc. We thus expect this migration to the peak to occur on a timescale  $(H/R)^2$  shorter than the timescale shown in Fig. 12. Simulations of a two component self-gravitating disc, where both a gaseous and a solid component

are present have been run [21, 22] and essentially confirm the scenario outlined above. Fig. 13 shows a comparison of the density structure of solids of different sizes, i.e. 10 meters (left) and 50 cm (right) [21]. Note that in these simulations the parameter  $\Omega\tau_{\text{cool}}$  was taken to be much larger than unity in order to prevent the gaseous disc to fragment. In the 10 meter case, the solid surface density essentially follows that of the gas, while in the 50 cm case, the solid concentrate in a filamentary structure at the peak of the spiral arm. This strong concentration can have important effects on the growth rate of solids in this size range. First of all, the enhanced density favours and speeds up collisional growth. Additionally (and possibly more importantly), the solid component might become gravitationally unstable, both because of the high density and because the velocity dispersion of solid particles is strongly reduced as the particles get ‘trapped’ within the spiral. Simulations of the same setup, and including the self-gravity of the solid component, have shown that meter sized solids in the filaments easily collapse and form gravitationally bound larger objects [22]. Unfortunately, the mass resolution of such simulations is too small to determine whether the resulting solid fragments would be of the right mass range for planetesimals.

**7.5.2. The gravitational instability model.** In the gravitational instability model gaseous giant planets are simply formed from the fragmentation of a self-gravitating gaseous disc [122-125]. As we have seen in section 4 above, this naturally follows from the development of gravitational instabilities, *if the cooling rate in the disc is sufficiently large*, such that  $\Omega\tau_{\text{cool}} < 1$ . If this happens, then gravitationally bound objects are able to form rapidly, within a few orbital timescales. With respect to the core accretion model, the gravitational instability model has the advantage of being a fast process, hence obviating to one of the difficulties associated with core accretion. A downside of it is that it can only account for the gas giant planets and would not explain the formation of terrestrial ones. Additionally, in this model gas giants would not have a solid core. Now, while for Jupiter this might also be the case, a solid core is thought to be present in Saturn. In a few cases, extra-solar planets have also been detected through eclipses of their parent stars [126], which allows to put constraints on their size and hence on their inner density, which can then give hints on the presence of a solid core. In one extreme case, HD 149026b, which has a total mass of  $0.36M_{\text{Jupiter}}$ , it is estimated that the core would be as massive as  $70M_{\oplus}$  [127], which would then exclude the gravitational instability model (but would also be very difficult for the core accretion one!).

What would be the typical mass of a planet formed by gravitational instability? A simple estimate can be obtained by considering that, in a thin disc, the most unstable wavelength is  $\lambda = 2\pi H$ . The typical mass of a fragment would then be of the order of  $M_{\text{frag}} \approx \Sigma\lambda^2$ . This can be simply rewritten in terms of the mass of the star  $M$  and the aspect ratio  $H/R$ , as

$$(73) \quad M_{\text{frag}} \approx \frac{4\pi}{Q} \left( \frac{H}{R} \right)^3 M.$$

Now, a gravitationally unstable disc has  $Q \approx 1$ , the typical aspect ratio of protostellar discs is  $H/R \approx 0.1$ , and if we take the central star mass  $M = M_{\odot}$ , we get  $M_{\text{frag}} \approx 12M_{\text{Jupiter}}$ . It then appears that this model would only form the most massive observed planets.

However, most of the debate over the gravitational instability model for planet formation concentrates on whether the actual cooling time in protostellar discs can be short

enough to induce fragmentation. This has been estimated analytically [128] and the results is that the cooling time is short enough to induce fragmentation only at very large distances from the central star, of the order of 50-100 AU. On the other hand, it has also been argued that if the disc becomes convectively unstable, then cooling might become more effective, producing the correct conditions to induce fragmentation ([129], but see [130] for a rebuttal of this hypothesis). Recent simulations that employ sophisticated (but still uncertain) radiative cooling prescriptions [51, 74] indicate that indeed fragmentation of protoplanetary discs only occurs at very large distances from the central star.

Even if the cooling time is long, it has been argued that the possible interaction with a companion or the close passage of a nearby star might trigger disc fragmentation by tidally increasing the local density. Early numerical isothermal simulations did indeed show such a behaviour [131-133]. However, it should be noticed that, as mentioned above, the response of an isothermal disc to a gravitational instability is bound to be relatively violent, since the stabilizing feedback loop associated with self-regulation cannot happen under isothermal conditions. More recently, the same issue has also been investigated with a non-isothermal equation of state and the results appear to be different. It has been found that both in the case of a disc in a binary system [134] and in the case of a stellar fly-by [135], the effect of tidal heating is much more effective in stabilizing the disc than the de-stabilizing effect of the associated density enhancement, so that even in this case the ultimate criterion for disc fragmentation is that the cooling time be shorter than the dynamical time (see [136] for an alternative view).

However, not all hope is lost for the gravitational instability model. As we have seen, this mechanism is indeed able to form large gaseous planets (or small brown dwarfs) at large radial distances from the host star, of the order of 100 AU. Such a new planetary population is still not excluded from observations. Note that the radial velocity technique requires tracking the stellar velocity for a period of the order of the planetary orbital period in order to detect a planet. Now, planets at very large distances have very long periods and might therefore require more time in order to be detected. Additionally, in the future such a different planet population might become detectable by other techniques, such as direct imaging.

In this respect, a very interesting case is that of the system 2MASS 1207. It consists of a brown dwarf primary, with a mass of  $25M_{\text{Jupiter}}$ , and a planetary mass companion, with a mass of  $5M_{\text{Jupiter}}$ , orbiting at a distance of 55 AU from the central brown dwarf [137] (recent revision of the distance to the system has provided slightly different orbital parameters [138], with revised masses of 21 and 3  $M_{\text{Jupiter}}$  for the primary and the secondary and a semi-major axis of 41 AU, but the following arguments remain essentially unchanged). The companion, 2MASS 1207b, is the first planetary mass object to be ever imaged. While clearly the absolute mass of the companion is in the planetary regime, it is questionable whether this system should be really considered as a planetary system around a brown dwarf or rather as an extremely low mass binary, since the mass ratio and the separation are perfectly consistent with the statistical properties of binaries of larger mass [139]. In any event, what is relevant here is to understand whether such a planetary mass object formed through the core accretion scenario or rather through gravitational instability. It has been shown [139] (see also [140]) that given the large separation of the system and the very low mass of the primary, it is extremely unlikely to form this object via core accretion. On the other hand, this is a good example where the gravitational instability scenario is expected to work, given the large mass ratio and the large separation. Whether such an object is a rare case, and especially whether cases

like this also occur around larger mass primaries, is still unknown but future observations might provide interesting constraints.

## 8. – APPLICATION TO OBSERVED SYSTEMS: AGN

After describing the impact of the disc self-gravity on the dynamics of accretion discs at the smallest scales, that is on the AU scale of protostellar and planet forming discs, we now turn to explore how it may affect the structure and dynamics of accretion discs at the largest scales, that is at the parsec scales of discs feeding supermassive black holes in Active Galactic Nuclei.

### 8.1. *Active Galactic Nuclei.* –

**8.1.1. Feeding the hole.** It is well known that accretion discs feed the growth of supermassive black holes in the nuclei of galaxies. Here the typical scale of the system is obviously much larger than in the case of protostellar discs. The central supermassive black hole has a mass that ranges from  $10^6$  to  $10^9 M_\odot$ . The radial extent of the disc is more uncertain. While the high energy emission from the AGN comes from the hotter innermost parts of the disc (from a few to a few tens Schwarzschild radii), there is also substantial evidence that the disc can extend much further out. One important piece of information in this context comes from the observation of water maser emission from AGN [2-4, 141], which reveals the presence of a thin disc of gas in almost Keplerian rotation out to a distance of the order of 1 pc, which corresponds to  $10^5 R_\bullet$  for a  $10^8 M_\odot$  black hole, where  $R_\bullet$  is the Schwarzschild radius. In some cases the kinematics of these maser spots presents some interesting features that will be discussed in more detail below. A very important difference between AGN discs and protostellar discs is that AGN discs are generally much thinner. Detailed models of the structure of the non-self-gravitating inner disc [142] indicate that the aspect ratio is of the order of  $H/R \approx 10^{-3}$ , although in the self-gravitating regime it can be larger, of the order of a few times  $10^{-2}$  [52]. This has important consequences for the issue of the role of the disc self-gravity. As we have often mentioned above, a simple way to check the importance of the disc self-gravity is to compare the mass ratio  $M_{\text{disc}}/M$  with the aspect ratio  $H/R$ . Now, given the small aspect ratio of AGN discs, a very light disc, with  $M_{\text{disc}}/M \approx 10^{-3}$  can already be subject to gravitational instabilities. It is then relatively easy to calculate the distance at which we would expect the cumulative disc mass to be high enough to make the disc unstable [19, 143]. It turns out that this distance is of the order of  $10^3 R_\bullet$ , or  $10^{-2}$  pc for a  $10^8 M_\odot$  black hole.

There is still some debate about the radial extent of the disc feeding the central black hole. Some models [144] postulate that the disc outer radius corresponds to the distance where it becomes self-gravitating, i.e.  $\approx 0.01 - 0.1$  pc, while some others allow the disc to extend out to  $10 - 100$  pc [145]. Probably reality falls somewhere in between. In fact, on the one hand observations of water maser emission clearly indicate the presence of a rotating gaseous disc at radii of the order of 1 pc, i.e. well within the self-gravitating portion of the disc. On the other hand it might be difficult to reconcile with observations discs which extend much further beyond this point. Let us see this in more details.

The main difference between between AGN discs and protostellar discs with respect to their behaviour in the self-gravitating regime lies in the different cooling rate, which for AGN is typically very short. This is usually seen by noting that the heating rate needed to keep the disc marginally stable at  $Q \approx 1$  is much larger than what can be provided by

a viscous accretion disc, with reasonable values of  $\alpha$  [143, 146]. Put it in other terms, this essentially means that the cooling timescale in AGN is much shorter than the dynamical timescale. If we now consider the results of the simulations described above in Section 4, we would thus conclude that the disc would fragment and form stars [147, 148]. This would be disastrous in terms of AGN feeding because star formation would essentially remove most of the fuel for accretion and turn it into stars. Indeed, such arguments are the main arguments in favour of AGN feeding from a small scale disc, but leave us with the problem of explaining the presence of large maser emitting discs.

One possible way out of this apparent paradox can be found if energy and angular momentum are transported by global spiral waves in the disc, rather than by a local  $\alpha$ -like process. As we have seen above, this usually happens if the disc mass is a sizable fraction of the central object mass. In this case, as already discussed, the energy balance should include some extra “global” terms [79], arising from wave transport of energy, that might provide the required energy to prevent fragmentation in the outer disc. In this picture, a density wave might remove free rotational energy from the inner disc, but rather than dissipating it locally (as would a standard viscous process do), it might carry it a long way out along the wave, and release it at large radii, where the wave is dissipated <sup>(2)</sup>. As seen above, this requires the presence of quite massive discs, and the evolution in this case is generally highly variable, with episodes of strong accretion and black hole feeding followed by more quiescent periods where the accretion rate is small. Such a time variable accretion model has also been sometimes proposed [149]. In some way, this solution is similar to the proposed solution for the high mass accretion rates observed around high mass young stars above (Section 7.4).

Alternatively, even if fragmentation does occur, it is still unclear whether accretion would be inhibited altogether, or whether we might still have accretion with only a relatively small amount of mass ending up in stars. Numerical simulations [150] show that while it is true that for small cooling times the disc undergoes fragmentation, if a small amount of energy is released from the forming stars (either from accretion luminosity from the stars or from the ignition of nuclear burning), this is sufficient to prevent further fragmentation (see also [151]). Depending on the parameters (and in particular if the cooling time is close to the fragmentation threshold) the lifetime of the gaseous disc can be very long. In this case, the missing energy to keep the disc self-regulated would be provided by star formation feedback rather than by global energy transport, as in the model above.

In both models described above, the disc is then allowed to extend outside  $\sim 10^3 R_\bullet$ , that is the radius where it becomes self-gravitating, provided that some extra-heating source is able to maintain it close to marginal stability in a self-regulated way. We have already noticed that self-regulated discs are thus characterized by a secondary bump in the broadband Spectral Energy Distribution, which can be prominent at infrared wavelengths. As shown in Fig. 4, the luminosity in this long wavelength bump is an increasing function of the location of the outer radius of the disc [49]. This fact can then be used to put constraints on the disc outer edge [52]. It turns out that if the disc extends beyond  $\approx 1$  pc, (and assuming that the self-regulation condition is determined

---

<sup>(2)</sup> Note that even a standard viscous process does transport energy between different regions of the disc, but this does not affect the arguments above, where we assume that in a wave-dominated transport the relative contributions of dissipation and transport of energy are significantly different from a viscous one.



by thermal pressure, rather than by turbulent motions of gas clumps and clouds within the disc) then the infrared bump would become too luminous, exceeding the typical observed SED of Active Galactic Nuclei. We thus see that, while an approximately self-regulated accretion disc can be present well beyond 0.01 pc, thus potentially explaining the presence of dense gaseous discs to produce the water maser emission, it is unlikely to extend far beyond 1 pc.

**8'1.2. Star formation in the Galactic Center.** In the context delineated above, an important role is played by the evidence that has been gathered in the last few years, which points to the presence of a large number of young stars very close to the supermassive black hole at the center of our own Milky Way [152,153]. In particular, most of these stars appear to belong to two distinct stellar discs orbiting at roughly the same distance to the black hole, i.e. at a distance of 0.05-0.5 pc [154,155]. The most likely explanation for the origin of these stars is that they formed *in situ* and in particular from the fragmentation of a self-gravitating accretion disc [154,156]. Such observations thus fit naturally in the context described above, since we know that at parsec distances an AGN accretion disc would be self-gravitating and its cooling time is expected to be short enough to induce fragmentation. The conditions in the Galactic Center might be typical of other galaxies, where a nuclear starburst can be a result of the very same mechanism [151,157].

**8'1.3. Kinematics of water masers.** In the discussion above, we have often referred to the detection of water maser emission from the inner parts of AGN. Indeed, such observations are a valuable tool to study the disc at the radial distances where self-gravity is expected to be important. In the section above, we used water maser observations simply to indicate the mere presence of discs at radii of the order of 0.1-1 pc. On the other hand they also offer a way to probe the disc geometry (for example, the presence of a warp [158-160]) and, most importantly, a way to probe its kinematic. In fact, water masers are usually observed to arise from a series of distinct “maser spots”, at different radial distances from the nucleus, and from the Doppler shift of the maser emission line it is then possible to reconstruct the rotation curve. In some cases, for example for NGC 4258, the resulting rotation curve is very close to Keplerian [2], and it thus allows a very precise determination of the mass of the central BH, which for NGC 4258 is  $3.6 \cdot 10^7 M_\odot$ .

However, in many other cases the rotation curve, while still displaying a smooth declining profile, as would be expected for a rotating disc, does not follow exactly Kepler's law. This is for example the case of NGC 1068 [3,141], of the Circinus galaxy [160] and of NGC 3079 [161]. In particular, for the case of NGC 1068 the maser data are consistent with a circular velocity  $v_\phi \propto r^{-0.31}$  [141]. Given the discussion above, which shows that at a scale of a fraction of a parsec, where the maser spots are detected, the disc can be self-gravitating, it is then tempting to attribute such (often small) deviation from Keplerian rotation to the contribution of the disc self-gravity.

A detailed fit to the circular velocity traced by water masers in NGC 1068 with a model which incorporates both the gravitational field of the black hole and that of the disc has been performed [19] (see also [162]), by using the self-regulated disc models described in Section 3'3. The results of such modeling are shown in Fig. 14, which illustrates the observed rotation curve and the fit with a self-regulated model [19]. As seen in Fig. 14, the observed rotation curve of the maser spots can be divided in two different regions: 1) at small impact parameter the masers show an apparently rising rotation curve; 2) starting from a radial distance  $\approx 0.6$  pc from the center the rotation

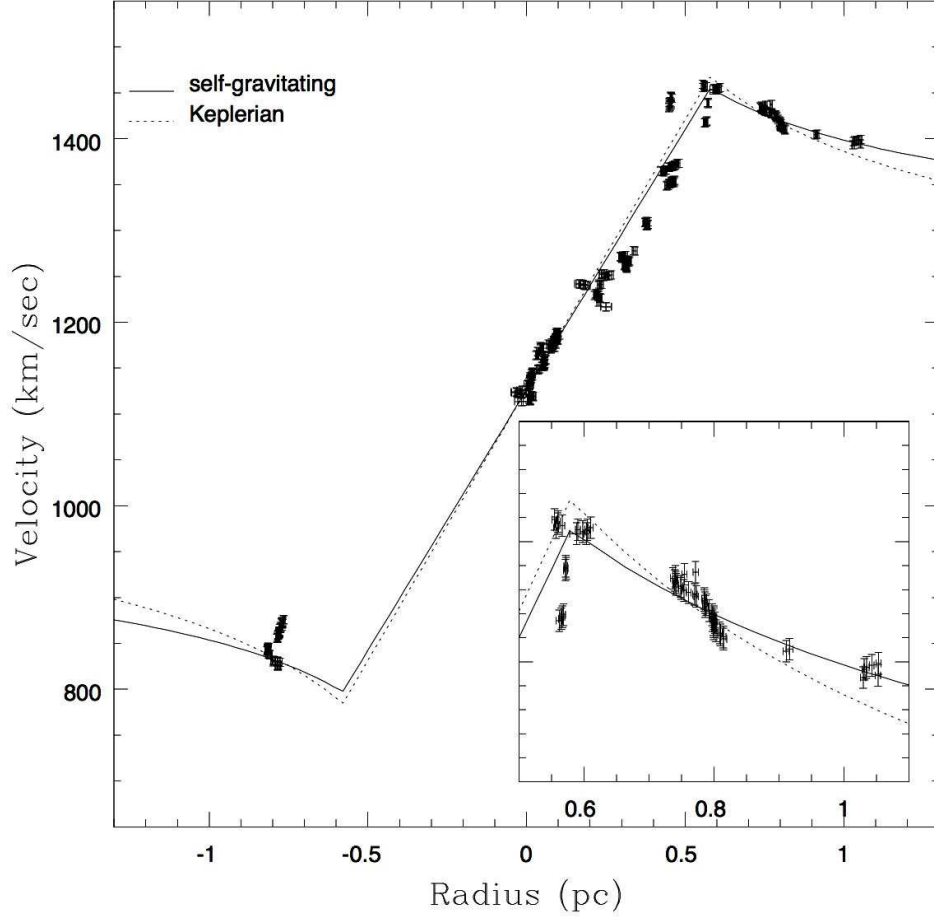


Fig. 14. – The data points show the observed rotation curve measured from water maser emission in the AGN NGC 1068. The solid line is the best-fit (non-Keplerian) model to the data, obtained from a self-regulated, self-gravitating disc model (see Section 3.3). In this model both the disc and the black hole contribute to the radial gravitational field, and we have  $M_{\text{disc}} \approx M = (8.0 \pm 0.3) 10^6 M_{\odot}$ . The dashed line is a fit obtained from a Keplerian model, where the mass of the black hole is  $M \approx (1.50 \pm 0.02) 10^7 M_{\odot}$ .

curve declines, following a sub-Keplerian profile.

The natural interpretation of the declining part of the rotation curve is that it arises from material that moves parallel to our line of sight (i.e. that lies on a disc diameter perpendicular to the line of sight). The best argument in favor of this interpretation is that maser amplification is largest for material that lies close to the line of the nodes. On the other hand, the rising part of the observed “rotation curve” is thought to originate from one quarter of the disc at the inner maser disc radius, so that the rising curve is an effect of velocity projection along the line of sight (see also [2]). According to this interpretation, the inner radius of the maser disc is located at  $\approx 0.6$  pc and the outer

disc radius is at  $\approx 1$  pc.

The best fitting self-gravitating model is shown in Fig. 14 with a solid line. There are essentially two parameters that determine the model, which can be traced back to the black hole mass  $M$  and the disc surface density  $\Sigma$ , which in turn (see Section 3.3) gives us information on  $\dot{M}/\alpha$ . The resulting black hole mass is  $M = (8.0 \pm 0.3) 10^6 M_\odot$  and the disc mass is approximately equal to the black hole mass. From the required disc surface density it is then possible to obtain  $\dot{M} = (28.1 \pm 0.2) \alpha M_\odot/\text{yr}$ . The mass accretion rate  $\dot{M}$  can be estimated for example from the bolometric luminosity as  $\dot{M} \approx 0.23 M_\odot/\text{yr}$ , and we thus obtain  $\alpha \approx 8.3 \cdot 10^{-3}$ , which is of the right order of magnitude as would be expected from the transport induced by gravitational instabilities.

The best-fit curve resulting from the self-gravitating disc model is not a power-law. However, as mentioned above the data are consistent with a rotation curve of the form  $V \propto r^{-0.31}$  [141]. Indeed, computing the quantity  $d \ln v_\phi / d \ln R$  for the best fit model, it is easily seen that it ranges from  $-0.35$  at the inner disc edge of the disc to  $-0.30$  at the outer disc edge.

The dashed line in Fig. 14 shows the result of a fit done with a Keplerian disc model. The fit is significantly worse than the non-Keplerian one (a detailed statistical analysis can be found in [19]). The resulting best-fit value of the black hole mass in the Keplerian fit is  $M \approx (1.50 \pm 0.02) 10^7 M_\odot$ . Note that the total mass (disc + black hole) of the self-gravitating best fit model is roughly the same as the black hole mass of the Keplerian fit. Therefore, a non-self-gravitating fit, which attributes all the mass to the central object, gives the correct value for the total mass, but fails to provide the correct slope of the rotation curve.

**8.2. Supermassive black hole formation.** – In the section above we have concentrated on supermassive black holes in the local Universe, i.e. in the Milky Way and in nearby AGN. We now turn our attention to very high redshifts, where we expect the emergence of the seeds of the first supermassive black holes. We will then see that even at such early epochs, phenomena related to the development of gravitational instabilities might play a key role in determining the properties of the black hole population.

Observations of AGN at high redshifts, up to  $z \sim 6$  [163, 164], indicate that supermassive black holes, with masses up to  $10^9 M_\odot$  were already in place when the Universe was only  $10^9$  years old. This clearly requires that the black hole growth occurred at very high rates, with an average of  $1 M_\odot/\text{yr}$ . Such a rapid early growth poses serious challenges to models of their formation. Some models [165–167] assume that the seeds of supermassive black holes are the remnants of the zero-metallicity first stars (the so-called Population III stars), that are expected to be relatively massive [168, 169] and thus produce black holes with a mass of up to  $100 M_\odot$ . However, unless the efficiency of conversion of matter into energy through the accretion process is very low, it is impossible to grow the seeds to the required masses by  $z \sim 6$  through Eddington-limited accretion [170]. The problem here is that when the accretion rate is large, the radiation pressure produced by the accretion luminosity can exceed the gravitational force of the black hole and hence lead to an expulsion instead of accretion of matter. The limiting luminosity at which this happens is called the Eddington limit. Now, if the accretion efficiency  $\epsilon = L/\dot{M}c^2$  exceeds  $\approx 0.1$  (where  $L$  is the accretion luminosity and  $c$  is the speed of light), the Eddington limit does not allow the large accretion rates needed to grow the seeds fast enough to become bright AGN by  $z \sim 6$  [170]. Note also that the Eddington limit is linearly proportional to the black hole mass, so that the problem of accreting at very high rates is particularly important in the earliest phases of the growth, when the black hole mass is

small. Once the black hole mass has exceeded  $10^4 - 10^5 M_\odot$ , further growth at  $\approx 1 M_\odot/\text{yr}$  can proceed at sub-Eddington rates.

The efficiency is in turn dependent on the spin of the black hole, with high spin producing very large efficiencies  $\epsilon \sim 0.5$ . Accretion of matter naturally tends to spin up the hole [166] and hence to increase the efficiency, thus exceeding the Eddington limit for relatively low  $\dot{M}$  and preventing a fast growth of the hole. While recent calculations [171, 172] show that it is possible to keep the hole spin low if the growth occurs through several small randomly oriented accretion episodes [173], we still have to face the issue of how to produce the high infall rates required.

Alternative models propose the direct formation of more massive seeds with masses of about  $10^5 M_\odot$  directly out of the collapse of dense gas [174-180]. The key limiting factor for these models is the disposal of the angular momentum. Recently, it has been proposed [179-182] that large scale gravitational instabilities developing during the growth of pre-galactic discs is the missing ingredient, able to funnel the required amount of gas into the center of the galaxy.

According to such models, the formation of the seeds of supermassive black holes occurs at a redshift  $z \sim 10 - 15$ , when the intergalactic medium had not been yet enriched by metals forming in the first stars. As a consequence, the chemical composition of the gas at this early epoch is essentially primordial, i.e. the gas is mostly hydrogen and helium. The cooling properties of this gas are therefore relatively simple. In particular, in the absence of molecular hydrogen, the main coolant is provided by atomic hydrogen, for which the cooling timescale becomes extremely long for temperatures smaller than  $\sim 10^4$  K, and we thus expect the gas to reach thermal equilibrium at a temperature  $T_{\text{gas}}$  of the order of  $10^4$  K.

Now, consider a dark matter halo (modeled, for simplicity, as a truncated singular isothermal sphere) of mass  $M_{\text{halo}}$  and circular velocity  $V_h$ , extending out to  $r_h = GM_{\text{halo}}/V_h^2$ . We also assume that the halo contains a gas mass  $M_{\text{gas}} = m_d M_{\text{halo}}$ , where  $m_d$  is of the order of the universal baryonic fraction,  $\approx 0.1$ , whose angular momentum is  $J_{\text{gas}} = j_d J$ , where  $j_d \sim m_d$ . The angular momentum of the dark matter halo  $J$  is expressed in terms of its spin parameter  $\lambda = J|E|^{1/2}/GM_{\text{halo}}^{5/2}$ , where  $E$  is its total energy. The probability distribution of the spin parameter of dark matter halos can be obtained from cosmological N-body simulations [183] and is well described by a log-normal distribution peaking at  $\lambda = 0.05$ .

If the virial temperature of the halo  $T_{\text{vir}} \propto V_h^2$  is larger than the gas temperature  $T_{\text{gas}}$ , the gas collapses and forms a rotationally supported disc, with circular velocity  $V_h$ , determined by the gravitational field of the halo. For low values of the spin parameter  $\lambda$  the resulting disc can be compact and dense. In this case, during the infall of gas onto the disc, its density rises until the stability parameter  $Q$  becomes of the order of unity. At this point, the disc starts developing a gravitational instability, which as we have seen above is able to efficiently redistribute angular momentum and allow accretion. Further infall of gas does not cause the density to rise much further, but rather it promotes an increasingly high accretion rate into the center. This process goes on until infall is over and the disc has attained a surface density low enough to be marginally gravitationally stable, i.e. with  $Q = \bar{Q}$ . It is then possible to calculate what fraction of the infalling mass needs to be transported into the center to make the disc marginally stable, as a

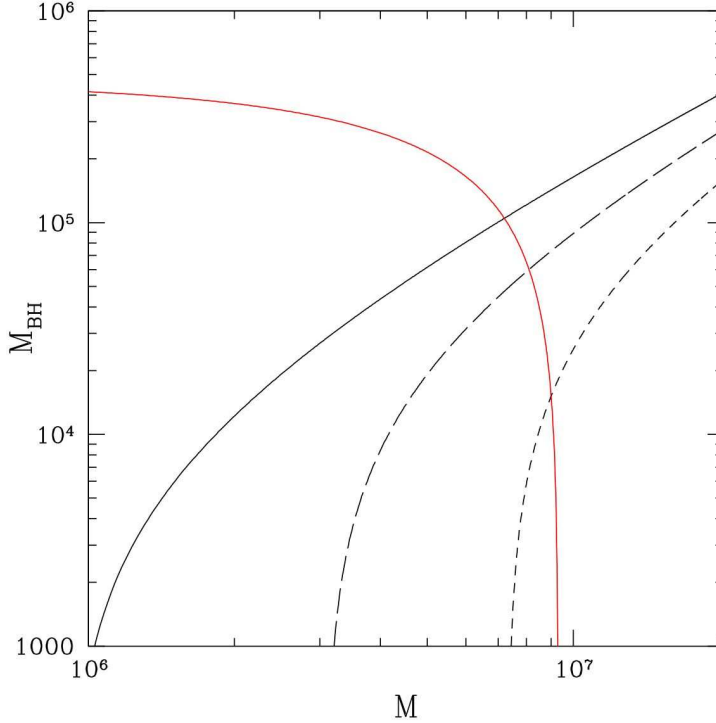


Fig. 15. – Mass available for the formation of the seed of a supermassive black hole in the center of pre-galactic discs as a function of the mass of the parent dark matter halo. The plots refer to the following choice of parameters:  $\bar{Q} = 2$ ,  $T_{\text{gas}} = 4000\text{K}$ ,  $m_d = j_d = 0.05$ ,  $\lambda = 0.01$  (solid line),  $\lambda = 0.015$  (long-dashed line),  $\lambda = 0.02$  (short-dashed line). The red curve shows the threshold for fragmentation from Equation (75), with  $\alpha_c = 0.06$ . Halos on the right of the red line give rise to fragmenting discs.

function of the main parameters involved. In this way, we get [181, 182]:

$$(74) \quad M_{\text{BH}} = m_d M_{\text{halo}} \left[ 1 - \sqrt{\frac{8\lambda}{m_d \bar{Q}} \left( \frac{j_d}{m_d} \right) \left( \frac{T_{\text{gas}}}{T_{\text{vir}}} \right)^{1/2}} \right],$$

where we have suggestively called  $M_{\text{BH}}$  the accreted mass, since this mass is the total mass available for the formation of the black hole seed in the center.

However, for large halo mass, the internal torques needed to redistribute the excess baryonic mass become too large to be sustained by the disc, which might then undergo fragmentation. We have seen in the previous sections that the maximum torque that can be delivered by a quasi-steady self-regulated disc is of the order of  $\alpha_c \approx 0.06$ . Since the infall rate of gas from the halo is proportional to  $T_{\text{vir}}^{3/2}$ , we expect fragmentation when the virial temperature exceeds a critical value  $T_{\text{max}}$ , given by (see [181] for details):

$$(75) \quad \frac{T_{\text{max}}}{T_{\text{gas}}} > \left( \frac{4\alpha_c}{m_d} \frac{1}{1 + M_{\text{BH}}/m_d M_{\text{halo}}} \right)^{2/3},$$

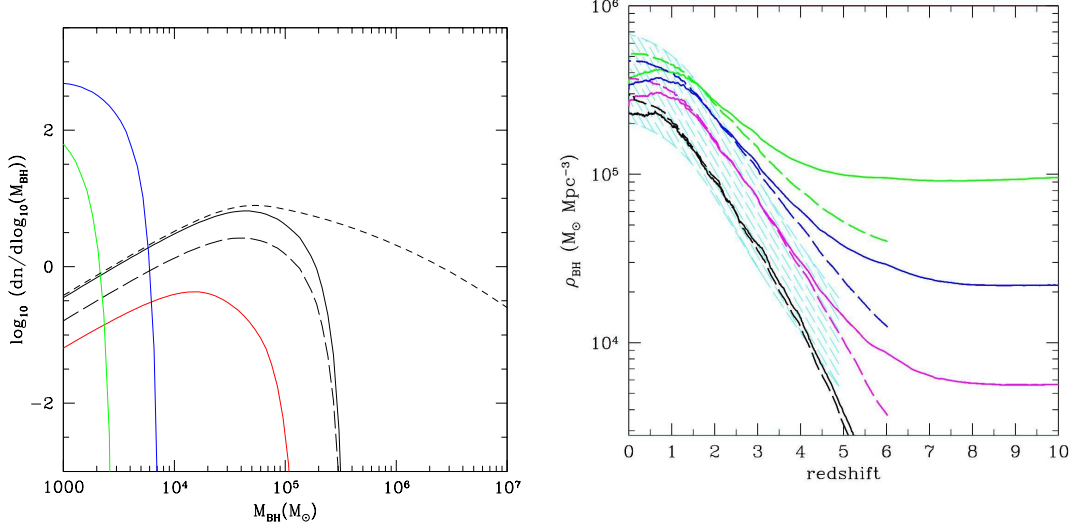


Fig. 16. – Left panel: Mass function of seed black holes predicted by the model based on Eqs. (74) and (75). The black solid line refers to  $z = 10$ , while the red line refers to  $z = 20$ . The long-dashed line shows the effect of reducing  $\bar{Q}$  from 2 to 1.5. The short dashed line shows the effect of not including the possibility of fragmentation (more details can be found in [182]). Right panel: the integrated density of black holes predicted from a merger tree evolution of the black hole seed population. The three solid line refer to three different choices of the parameters (see [184] for details). The dashed area indicate the observationally permitted region from estimates of the black hole density at low redshifts. At high redshift ( $z \gtrsim 6$ ) the density reflects the one attained after the seed formation phase, while the rise at lower redshifts indicates the growth through AGN activity.

Although it is possible, as often mentioned in this paper, that accretion proceeds even for larger values of  $\alpha$  in a highly time-variable way when the disc mass is large, and it is also possible that accretion proceeds even in a fragmenting disc, we make here the conservative assumption that all halos that violate Eq. (75) do fragment and do not accrete. Fig. 15 illustrates the relationship between halo mass and black hole mass based on Equation (74) for three different values of the spin parameter  $\lambda$ . The red line in Fig. 15 correspond to Eq. (75), so that halos on the right of the red line are expected to fragment. We can thus see that the typical mass fed into the center of such pre-galactic disc is of the order of  $10^3 M_{\odot}$  up to  $10^5 M_{\odot}$ . The typical accretion rates during this early epochs is of the order of  $10^{-2} M_{\odot}/\text{yr}$  [181]. If such high masses are assembled as seeds of supermassive black holes at redshift 10 – 15, it is then easy to grow through Eddington-limited accretion to  $10^9 M_{\odot}$  by  $z = 6$ , as required by observations.

Equation (74) provides a powerful link between the properties of dark matter haloes and the mass of massive seed black holes that can grow within them. As shown, the amount of mass that will be concentrated in the central regions of these pre-galactic discs depends only on halo properties (such as the spin parameter  $\lambda$  and the fraction of baryonic mass that collapses to the disc  $m_d$ ), on the ratio between gas temperature and halo virial temperature, and on the threshold value of  $Q$ , which has a very small range of variation around  $\bar{Q} \approx 1$ . This simple model has been used to calculate several properties

of the black hole population at high redshift. In particular, from the distribution of halo masses and angular momentum it is straightforward to derive the mass function of the supermassive black hole seeds [182], which turns out to be strongly peaked at around  $10^5 M_\odot$ , as shown in the left panel of Fig. 16. Furthermore, it is also possible to include such a simple prescription within evolutionary models that track the properties of the black hole population along cosmic time, such as merger tree models [184] (such models are usually based on the so called hierarchical model for the growth of structures, where the structures observed in the present day Universe are the result of the subsequent merger of smaller structures at early times). It is then interesting to see that the evolution of such a primordial seed population can naturally account for the current estimates of the density of black holes at low redshift (Fig. 16, right panel). In addition, an important and testable prediction of such models is that dwarf galaxies, that did not have any progenitor massive enough to seed a black hole, should not host a supermassive black hole. In particular if the velocity dispersion of the galaxy is below  $\sim 50$  km/sec, the probability of hosting a black hole turns out to be negligibly small [184].

## 9. – CONCLUSIONS AND CHALLENGES FOR THE FUTURE

As we have seen, self-gravitating accretion discs are an important aspect of the modeling of several systems, and thus play a key role in modern astrophysics. The influence of the disc self-gravity extends from the very small scales, associated with the formation of planetary systems, up to the large scales, associated with the formation and the feeding of supermassive black holes in the nuclei of galaxies. While clearly the specific properties of systems associated with such hugely different scales are going to differ substantially, there is yet a unifying framework related to the scale-free nature of the gravitational force.

Because of self-gravity, accretion discs can develop gravitational instabilities. These are going to deeply affect the structure and the evolution of the accretion disc. Firstly, we have seen that the instability usually takes the form of a regular spiral structure, that can efficiently redistribute the disc angular momentum and thus allow the accretion process to take place. In this respect the role of the disc self-gravity is complementary to that of other instabilities, such as magneto-hydrodynamic (MHD) instabilities [31, 78]. Indeed, while MHD instabilities are likely to be the most important source of angular momentum transport in the inner and hotter parts of the disc, it is not clear whether they are able to provide the necessary torques in colder environments, where the ionization level of the gas is low and the coupling to the magnetic field is weaker. This is, for example, the case of the outer parts of protostellar discs [185]. On the other hand, such cold parts of the discs are naturally subject to gravitational instabilities. It is then possible to envisage a situation where gravitational instabilities are responsible for bringing the accreting material from large distances (which might be of the order of tens of AU for protostellar discs, and a fraction of a parsec for AGN discs) into the inner disc, where MHD instabilities take over and are responsible for the ultimate deposition of the matter onto the central object. In this respect, it would be important to develop models which include both the effects of MHD induced turbulence and that of gravitational instabilities. Semi-analytical models along these lines have been discussed [34], but it is still unclear what would be the interplay between magnetic and gravitational instabilities in the general case [186, 187].

A second aspect related to gravitational instabilities is the natural tendency associated with self-gravity to collapse and produce gravitationally bound objects. Disc fragmenta-

tion is a possible outcome of the onset of gravitational instabilities. As discussed above, it is now well established that the conditions leading to fragmentation are essentially related to the disc thermodynamics and to the cooling rate. From the modeling point of view, fragmentation has a twofold aspect. On the one hand, if one is interested in producing smaller objects within the disc, it offers a natural way to accomplish the task. On the other hand, there is the danger that most of the disc mass would end up in the fragments, with very little being accreted onto the central object. In this respect, fragmentation is a dangerous outcome, which might inhibit accretion.

We have seen that the biggest difference between protostellar discs and AGN discs with respect to self-gravity lies indeed in their thermal properties, and in particular in the fact that AGN discs are expected to be cold (in the sense that  $H/R \ll 1$ ) and with a short cooling timescale (in the sense that  $\tau_{\text{cool}}\Omega \ll 1$ ), while protostellar discs are hot ( $H/R \approx 0.1$ ) and with a long cooling timescale ( $\tau_{\text{cool}}\Omega \gg 1$ , at least within 100 AU). The behaviour of these two kind of systems in this respect is then significantly different. Most protostellar discs are not expected to fragment, except perhaps in their outermost regions. This is the main problem affecting models of planet formation which rely on direct disc fragmentation. At the other end of the scales that we have considered, we see that AGN discs instead are likely to fragment. This is good news in view of forming young massive stars in the vicinity of the central supermassive black hole, as are observed, for example, close to the black hole in our Galactic Center. On the other hand, we have discussed above the problems that might arise from disc fragmentation in this context, related to the feeding of the black hole.

The last 10-15 years have witnessed a significant improvement of our understanding of the evolution of self-gravitating accretion discs. This has been made possible through the interplay between simple analytical models and the results of detailed numerical simulations of the disc hydrodynamics. The big advances in computing power in the last decade have made it possible to run such numerical simulations with unprecedented resolution, and progressively including the effects of several physical processes in more detail. In this paper, emphasis has been put on the important role of the gas cooling rate in determining the outcome of the instability. This has been often introduced in a simplified way in the simulations, which has led to a valuable investigation of the different regimes, in a somewhat academic way. The theoretical challenge in this respect is to introduce in the numerical codes a more realistic cooling function, in order to understand the behaviour of actual systems. For example, most of the current debate on the applicability of the fragmentation scenario for planet formation is ultimately due to our ignorance of the actual thermodynamics at work in protostellar discs. The difficulty here is to treat numerically the radiative transfer within very optically thick media. Only very recently have numerical codes implemented radiation physics and, although still preliminary, some results are starting to emerge [51, 74]. Codes which couple radiative transfer with the disc hydrodynamics are also desired to include the possibly important effect of irradiation from the central object, which has mostly been neglected in the simulations run so far (with some exceptions [188]).

If we look at the problem of star formation within AGN accretion discs, here the issue is ultimately whether accretion can proceed at significant rates even if the disc fragments. In other words, it is still unclear how much gas needs to be turned into stars before fragmentation is quenched. Here, it would be important to include in the models the energy feedback arising from the forming stars. Even in this context then, the key ingredient to be added into numerical codes is a proper treatment of radiative transfer, coupled with hydrodynamics.



\* \* \*

ACKNOWLEDGMENTS. I would like to dedicate this paper to the memory of Eduardo Delgado-Donate, who contributed to the development of some of the ideas presented here. I acknowledge several interesting discussions with Giuseppe Bertin, Cathie Clarke, Andrew King, Sergei Nayakshin and Jim Pringle. I would also like to thank the various collaborators with whom I have worked on these issues along the years, and in particular Phil Armitage, Priya Natarajan and Ken Rice.

## REFERENCES

- [1] BERTIN G. and LIN C. C., *Spiral Structure in Galaxies: a Density Wave Theory* (MIT Press, Cambridge) 1996.
- [2] MIYOSHI M. *et al.*, *Nature* , **373** (1995) 127.
- [3] GREENHILL L. J. and GWINN C. R., *Astrophysics & Space Science* , **248** (1997) 261.
- [4] KONDRATKO P. T., GREENHILL L. J. and MORAN J. M., *ApJ* , **652** (2006) 136.
- [5] FRANK J., KING A. and RAINE D., *Accretion Power in Astrophysics* (Cambridge University Press, Cambridge) 2002.
- [6] REIPURTH B., JEWITT D. and KEIL K. (Eds.), *Protostars and Planets V* (Arizona University Press) 2007.
- [7] PACZYŃSKI B., *Acta Astronomica* , **28** (1978) 91.
- [8] KOLYKHALOV P. I. and SUNYAEV R. A., *Soviet Astronomy Letters* , **5** (1979) 180.
- [9] LIN D. N. C. and PRINGLE J. E., *MNRAS* , **225** (1987) 607.
- [10] LIN D. N. C. and PRINGLE J. E., *ApJ* , **358** (1990) 515.
- [11] PRINGLE J. E., *ARA&A* , **19** (1981) 137.
- [12] BINNEY J. and TREMAINE S., *Galactic dynamics* (Princeton, NJ, Princeton University Press, 1987, 747 p.) 1987.
- [13] BERTIN G. and LODATO G., *A&A* , **350** (1999) 694.
- [14] MESTEL L., *MNRAS* , **126** (1963) 553.
- [15] BERTIN G., *ApJ* , **478** (1997) L71.
- [16] BECKWITH S. V. W., SARGENT A. I., CHINI R. S. and GUESTEN R., *AJ* , **99** (1990) 924.
- [17] CESARONI R. *et al.*, *ApJ* , **435** (1994) 137.
- [18] CESARONI R., NERI R., OLM I., TESTI L., WALMSLEY C. M. and HOFNER P., *A&A* , **434** (2005) 1039.
- [19] LODATO G. and BERTIN G., *A&A* , **398** (2003) 517.
- [20] WEIDENSCHILLING S., *MNRAS* , **180** (1977) 57.
- [21] RICE W. K. M., LODATO G., PRINGLE J. E., ARMITAGE P. J. and BONNELL I. A., *MNRAS* , **355** (2004) 543.
- [22] RICE W. K. M., LODATO G., PRINGLE J. E., ARMITAGE P. J. and BONNELL I. A., *MNRAS* , **372** (2006) L9.
- [23] SPITZER L., *ApJ* , **95** (1942) 329.
- [24] TOOMRE A., *ApJ* , **139** (1964) 1217.
- [25] BERTIN G., *Dynamics of Galaxies* (Cambridge University Press, Cambridge) 2000.
- [26] SHAKURA N. I. and SUNYAEV R. A., *A&A* , **24** (1973) 337.
- [27] HARTMANN L., CALVET N., GULLBRING E. and D’ALESSIO P., *ApJ* , **495** (1998) 385.
- [28] LASOTA J. P., *New Astronomy Reviews* , **45** (2001) 449.
- [29] BELL K. R. and LIN D. N. C., *ApJ* , **427** (1994) 987.
- [30] LODATO G. and CLARKE C. J., *MNRAS* , **353** (2004) 841.
- [31] BALBUS S. A. and HAWLEY J. F., *Reviews of Modern Physics* , **70** (1998) 1.
- [32] HARTMANN L., D’ALESSIO P., CALVET N. and MUZEROLLE J., *ApJ* , **648** (2006) 484.
- [33] CLARKE C. J., *MNRAS* , **376** (2007) 1350.

- [34] KRATTER K. M., MATZNER C. D. and KRUMHOLZ M. R., *ArXiv e-prints* , **0709.4252** (2007) .
- [35] LYNDEN-BELL D. and PRINGLE J. E., *MNRAS* , **168** (1974) 603.
- [36] HARTMANN L., *Accretion Processes in Star Formation* (Cambridge University Press, Cambridge) 1998.
- [37] LYNDEN-BELL D., *Nature* , **223** (1969) 690.
- [38] ADAMS F. C., LADA C. J. and SHU F. H., *ApJ* , **326** (1988) 865.
- [39] CHIANG E. P. and GOLDREICH P., *ApJ* , **490** (1997) 368.
- [40] DULLEMOND C. P., HOLLENBACH D., KAMP I. and D’ALESSIO P., *Protostars and Planets V* , (2007) 555.
- [41] HOHL F., *ApJ* , **168** (1971) 343.
- [42] OSTRICKER J. P. and PEEBLES P. J. E., *ApJ* , **186** (1973) 467.
- [43] LAU Y. Y. and BERTIN G., *ApJ* , **226** (1978) 508.
- [44] BERTIN G., LIN C. C., LOWE S. A. and THURSTANS R. P., *ApJ* , **338** (1989) 104.
- [45] LODATO G. and RICE W. K. M., *MNRAS* , **351** (2004) 630.
- [46] LODATO G. and RICE W. K. M., *MNRAS* , **358** (2005) 1489.
- [47] VANDERVOORT P. O., *ApJ* , **161** (1970) 87.
- [48] BERTIN G. and ROMEO A. B., *A&A* , **195** (1988) 105.
- [49] LODATO G. and BERTIN G., *A&A* , **375** (2001) 455.
- [50] VOROBYOV E. I. and BASU S., *MNRAS* , (2007) 855.
- [51] BOLEY A. C., MEJÍA A. C., DURISEN R. H., CAI K., PICKETT M. K. and D’ALESSIO P., *ApJ* , **651** (2006) 517.
- [52] SIRKO E. and GOODMAN J., *MNRAS* , **341** (2003) 501.
- [53] MILLER R. H., PRENDERGAST K. H. and QUIRK W. J., *ApJ* , **161** (1970) 903.
- [54] HOHL F., *ApJ* , **184** (1973) 353.
- [55] ANTHONY D. M. and CARLBERG R. G., *ApJ* , **332** (1988) 637.
- [56] LAUGHLIN G. and BODENHEIMER P., *ApJ* , **436** (1994) 335.
- [57] LAUGHLIN G. and RÓŻYCKA M., *ApJ* , **456** (1996) 279.
- [58] LAUGHLIN G., KORCHAGIN V. and ADAMS F. C., *ApJ* , **504** (1998) 945.
- [59] PICKETT B. K. *et al.*, *ApJ* , **504** (1998) 468.
- [60] PICKETT B. K. *et al.*, *ApJ* , **529** (2000) 1034.
- [61] GAMMIE C. F., *ApJ* , **553** (2001) 174.
- [62] RICE W. K. M., ARMITAGE P. J., BATE M. R. and BONNELL I. A., *MNRAS* , **338** (2003) 227.
- [63] RICE W. K. M., ARMITAGE P. J., BATE M. R., BONNELL I. A., JEFFERS S. V. and VINE S. G., *MNRAS* , **346** (2003) L36.
- [64] RICE W. K. M., ARMITAGE P. J., BATE M. R. and BONNELL I. A., *MNRAS* , **339** (2003) 1025.
- [65] RICE W. K. M., LODATO G. and ARMITAGE P. J., *MNRAS* , **364** (2005) L56.
- [66] PICKETT B. K., MEJÍA A. C., DURISEN R. H., CASSEN P. M., BERRY D. K. and LINK R. P., *ApJ* , **590** (2003) 1060.
- [67] MEJIA A. C., DURISEN R. H., PICKETT M. K. and CAI K., *ApJ* , **619** (2005) 1098.
- [68] MAYER L., QUINN T., WADSLEY J. and STADEL J., *ApJ* , **609** (2004) 1045.
- [69] REES M. J., *MNRAS* , **176** (1976) 483.
- [70] SILK J., *ApJ* , **214** (1977) 152.
- [71] CLARKE C., HARPER-CLARK E. and LODATO G., *MNRAS* , **381** (2007) 1543.
- [72] JOHNSON B. M. and GAMMIE C. F., *ApJ* , **597** (2003) 131.
- [73] WHITEHOUSE S. C., BATE M. R. and MONAGHAN J. J., *MNRAS* , **364** (2005) 1367.
- [74] MAYER L., LUFKIN G., QUINN T. and WADSLEY J., *ApJ* , **661** (2007) L77.
- [75] STAMATELLOS D., WHITWORTH A. P., BISBAS T. and GOODWIN S., *A&A* , **475** (2007) 37.
- [76] SHU F. H., *ApJ* , **160** (1970) 99.
- [77] LYNDEN-BELL D. and KALNAJS A. J., *MNRAS* , **157** (1972) 1.
- [78] BALBUS S. A., *ARA&A* , **41** (2003) 555.
- [79] BALBUS S. A. and PAPALOIZOU J. C. B., *ApJ* , **521** (1999) 650.

- [80] MCKEE C. F. and OSTRICKER E. C., *ARA&A* , **45** (2007) 565.
- [81] LADA C. J. and WILKING B. A., *ApJ* , **287** (1984) 610.
- [82] ANDREWS S. M. and WILLIAMS J. P., *ApJ* , **631** (2005) 1134.
- [83] ANDREWS S. M. and WILLIAMS J. P., *ApJ* , **659** (2007) 705.
- [84] EISNER J. A. and CARPENTER J. M., *ApJ* , **641** (2006) 1162.
- [85] TESTI L. *et al.*, *ApJ* , **554** (2001) 1087.
- [86] NATTA A., TESTI L., NERI R., SHEPHERD D. S. and WILNER D. J., *A&A* , **416** (2004) 179.
- [87] GULLBRING E., HARTMANN L., BRICEÑO C. and CALVET N., *ApJ* , **492** (1998) 323.
- [88] GULLBRING E., CALVET N., MUZEROLLE J. and HARTMANN L., *ApJ* , **544** (2000) 927.
- [89] ARTYMOWICZ P. and LUBOW S. H., *ApJ* , **467** (1996) 77L.
- [90] LIN S.-Y. *et al.*, *ApJ* , **645** (2006) 1297.
- [91] PIÉTU V., GUILLLOTEAU S. and DUTREY A., *A&A* , **443** (2005) 945.
- [92] RODRIGUEZ L. F., LOINARD L., D’ALESSIO P., WILNER D. and P.T.P. H., *ApJ* , **621** (2005) L133.
- [93] CLARKE C., LODATO G., MELNIKOV S. Y. and IBRAHIMOV M. A., *MNRAS* , **361** (2005) 942.
- [94] KENYON S. J. and HARTMANN L., *ApJ* , **383** (1991) 664.
- [95] MILLAN-GABET R. *et al.*, *ApJ* , **641** (2006) 547.
- [96] HARTMANN L. and KENYON S. J., *ApJ* , **299** (1985) 462.
- [97] KENYON S. J., HARTMANN L. and HEWETT R., *ApJ* , **325** (1988) 231.
- [98] LODATO G. and BERTIN G., *A&A* , **408** (2003) 1015.
- [99] BELL K. R., LIN D. N. C., HARTMANN L. W. and KENYON S. J., *ApJ* , **444** (1995) 376.
- [100] CLARKE C. J., LIN D. N. C. and PRINGLE J. E., *MNRAS* , **242** (1990) 439.
- [101] VOROBYOV E. I. and BASU S., *ApJ* , **633** (2005) L137.
- [102] VOROBYOV E. I. and BASU S., *ApJ* , **650** (2006) 956.
- [103] ARMITAGE P. J., LIVIO M. and PRINGLE J. E., *MNRAS* , **324** (2001) 705.
- [104] PALLA F. and STAHLER S. W., *ApJ* , **375** (1993) 288.
- [105] CESARONI R., GALLI D., LODATO G., WALMSLEY C. M. and ZHANG Q., in proc. of *Protostars and Planets V*, edited by REIPURTH B., JEWITT D. and KEIL K. 2007, p. 197.
- [106] CHINI R. *et al.*, *Nature* , **429** (2004) 155.
- [107] TORRELLES J. M., GÓMEZ J. F., GARAY G., RODRÍGUEZ L. F., CURIEL S., COHEN R. J. and HO P. T. P., *ApJ* , **509** (1998) 262.
- [108] GREENHILL L. J., GEZARI D. Y., DANCHI W. C., NAJITA J., MONNIER J. D. and TUTHILL P. G., *ApJ* , **605** (2004) L57.
- [109] PATEL N. A., CURIEL S., SRIDHARAN T. K., ZHANG Q., HUNTER T. R., HO P. T. P., TORRELLES J. M., MORAN J. M., GÓMEZ J. F. and ANGLADA G., *Nature* , **437** (2005) 109.
- [110] EDRIS K. A., FULLER G. A., COHEN R. J. and ETOKA S., *A&A* , **434** (2005) 213.
- [111] MAYOR M. and QUELOZ D., *Nature* , **378** (1995) 355.
- [112] UDRY S., BONFILS X., DELFOSSE X., FORVEILLE T., MAYOR M., PERRIER C., BOUCHY F., LOVIS C., PEPE F., QUELOZ D. and BERTAUX J.-L., *A&A* , **469** (2007) L43.
- [113] GUILLOT T., *Planetary and Space Sciences* , **47** (1999) 10.
- [114] POLLACK J. B. *et al.*, *Icarus* , **124** (1996) 62.
- [115] LISSAUER J. J. and STEVENSON D. J., in proc. of *Protostars and Planets V*, edited by REIPURTH B., JEWITT D. and KEIL K. 2007, p. 591.
- [116] ARMITAGE P. J., *ArXiv Astrophysics e-prints* , **0701485** (2007) .
- [117] VAN BOEKEL R. *et al.*, *A&A* , **437** (2005) 189.
- [118] APAI D., PASCUCCI I., BOUWMAN J., NATTA A., HENNING T. and DULLEMOND C. P., *Science* , **310** (2005) 834.
- [119] BERTIN G. and CAVA A., *A&A* , **459** (2006) 333.
- [120] HAGHIGHIPOUR N. and BOSS A. P., *ApJ* , **583** (2003) 996.
- [121] HAGHIGHIPOUR N. and BOSS A. P., *ApJ* , **598** (2003) 1301.
- [122] BOSS A. P., *ApJ* , **536** (2000) L101.

- [123] BOSS A. P., *ApJ* , **576** (2002) 462.
- [124] MAYER L., QUINN T., WADSLEY J. and STADEL J., *Science* , **298** (2002) 1756.
- [125] DURISEN R. H., BOSS A. P., MAYER L., NELSON A. F., QUINN T. and RICE W. K. M., in proc. of *Protostars and Planets V*, edited by REIPURTH B., JEWITT D. and KEIL K. 2007, p. 607.
- [126] CHARBONNEAU D., BROWN T. M., BURROWS A. and LAUGHLIN G., in proc. of *Protostars and Planets V*, edited by REIPURTH B., JEWITT D. and KEIL K. 2007, p. 701.
- [127] FORTNEY J. J., SAUMON D., MARLEY M. S., LODDERS K. and FREEDMAN R. S., *ApJ* , **642** (2006) 495.
- [128] RAFIKOV R., *ApJ* , **621** (2005) 69.
- [129] BOSS A. P., *ApJ* , **610** (2004) 456.
- [130] RAFIKOV R. R., *ApJ* , **662** (2007) 642.
- [131] BOFFIN H. M. J., WATKINS S. J., BHATTAL A. S., FRANCIS N. and WHITWORTH A. P., *MNRAS* , **300** (1998) 1189.
- [132] WATKINS S. J., BHATTAL A. S., BOFFIN H. M. J., FRANCIS N. and WHITWORTH A. P., *MNRAS* , **300** (1998) 1205.
- [133] WATKINS S. J., BHATTAL A. S., BOFFIN H. M. J., FRANCIS N. and WHITWORTH A. P., *MNRAS* , **300** (1998) 1214.
- [134] MAYER L., WADSLEY J., QUINN T. and STADEL J., *MNRAS* , **363** (2005) 641.
- [135] LODATO G., MERU F., CLARKE C. J. and RICE W. K. M., *MNRAS* , **374** (2007) 590.
- [136] BOSS A. P., *ApJ* , **641** (2006) 1148.
- [137] CHAUVIN G., LAGRANGE A.-M., DUMAS C., ZUCKERMAN B., MOUILLET D., SONG I., BEUZIT J.-L. and LOWRANCE P., *A&A* , **425** (2004) L29.
- [138] MAMAJEK E. E., *ApJ* , **634** (2005) 138.
- [139] LODATO G., DELGADO-DONATE E. and CLARKE C. J., *MNRAS* , **364** (2005) L91.
- [140] PAYNE M. J. and LODATO G., *MNRAS* , **381** (2007) 1597.
- [141] GREENHILL L. J., GWINN C. R., ANTONUCCI R. and BARVAINIS R., *ApJ* , **472** (1996) L21.
- [142] COLLIN-SOUFFRIN S. and DUMONT A. M., *A&A* , **229** (1990) 292.
- [143] GOODMAN J., *MNRAS* , **339** (2003) 937.
- [144] KING A. R. and PRINGLE J. E., *MNRAS* , **377** (2007) L25.
- [145] THOMPSON T. A., QUATAERT E. and MURRAY N., *ApJ* , **630** (2005) 167.
- [146] BERTIN G. and LODATO G., *A&A* , **370** (2001) 342.
- [147] SHLOSMAN I., FRANK J. and BEGELMAN M. C., *Nature* , **338** (1989) 45.
- [148] COLLIN S. and ZAHN J. P., *A&A* , **344** (1999) 433.
- [149] COLLIN S. and ZAHN J.-P., *ArXiv e-prints* , **0709.3772** (2007) .
- [150] NAYAKSHIN S., CUADRA J. and SPRINGEL V., *MNRAS* , **379** (2007) 21.
- [151] NAYAKSHIN S., *MNRAS* , **372** (2006) 143.
- [152] GENZEL R. *et al.*, *ApJ* , **594** (2003) 812.
- [153] GHEZ A. M. *et al.*, *ApJ* , **620** (2005) 744.
- [154] LEVIN Y. and BELOBORODOV A. M., *ApJ* , **590** (2003) L33.
- [155] PAUMARD T., GENZEL R., MARTINS F., NAYAKSHIN S., BELOBORODOV A. M., LEVIN Y., TRIPPE S., EISENHAEUER F., OTT T., GILLESSEN S., ABUTER R., CUADRA J., ALEXANDER T. and STERNBERG A., *ApJ* , **643** (2006) 1011.
- [156] NAYAKSHIN S. and SUNYAEV R., *MNRAS* , **364** (2005) L23.
- [157] LEVIN Y., *MNRAS* , **374** (2007) 515.
- [158] HERRNSTEIN J. R., GREENHILL L. J. and MORAN J. M., *ApJ* , **468** (1996) L17.
- [159] PAPALOIZOU J. C. B., TERQUEM C. and LIN D. N. C., *ApJ* , **497** (1998) 212.
- [160] GREENHILL L. J., KONDRATKO P. T., LOVELL J. E. J., KUIPER T. B. H., MORAN J. M., JAUNCEY D. L. and BAINES G. P., *ApJ* , **582** (2003) L11.
- [161] KONDRATKO P. T., GREENHILL L. J. and MORAN J. M., *ApJ* , **618** (2005) 618.
- [162] HURÉ J.-M., *A&A* , **395** (2002) L21.
- [163] FAN X. *et al.*, *AJ* , **128** (2004) 515.
- [164] FAN X. *et al.*, *AJ* , **131** (2006) 1203.
- [165] HAIMAN Z. and LOEB A., *ApJ* , **503** (1998) 505.

- [166] VOLONTERI M., MADAU P., QUATAERT E. and REES M. J., *ApJ*, **620** (2005) 69.
- [167] WYITHE J. S. B. and LOEB A., *ApJ*, **634** (2005) 910.
- [168] ABEL T., BRYAN G. L. and NORMAN M. L., *ApJ*, **540** (2000) 39.
- [169] BROMM V., COPPI P. S. and LARSON R. B., *ApJ*, **564** (2002) 23.
- [170] VOLONTERI M. and REES M. J., *ApJ*, **633** (2005) 624.
- [171] KING A. R., LUBOW S. H., OGILVIE G. I. and PRINGLE J. E., *MNRAS*, **363** (2005) 49.
- [172] LODATO G. and PRINGLE J. E., *MNRAS*, **368** (2006) 1196.
- [173] KING A. R. and PRINGLE J. E., *MNRAS*, **373** (2006) L90.
- [174] HAEHNELT M. G. and REES M. J., *MNRAS*, **263** (1993) 168.
- [175] UMEMURA M., LOEB A. and TURNER E. L., *ApJ*, **419** (1993) 459.
- [176] LOEB A. and RASIO F. A., *ApJ*, **432** (1994) 52.
- [177] EISENSTEIN D. J. and LOEB A., *ApJ*, **443** (1995) 11.
- [178] BROMM V. and LOEB A., *ApJ*, **596** (2003) 34.
- [179] KOUSHIAPPAS S. M., BULLOCK J. S. and DEKEL A., *MNRAS*, **354** (2004) 292.
- [180] BEGELMAN M. C., VOLONTERI M. and REES M. J., *MNRAS*, **370** (2006) 289.
- [181] LODATO G. and NATARAJAN P., *MNRAS*, **371** (2006) 1813.
- [182] LODATO G. and NATARAJAN P., *MNRAS*, **377** (2007) L64.
- [183] WARREN M. S., QUINN P. J., SALMON J. K. and ZUREK W. H., *ApJ*, **399** (1992) 405.
- [184] VOLONTERI M., LODATO G. and NATARAJAN P., *ArXiv e-prints*, **0709.0529** (2007) .
- [185] FROMANG S., TERQUEM C. and BALBUS S. A., *MNRAS*, **329** (2002) 18.
- [186] FROMANG S., BALBUS S. A. and DE VILLIERS J.-P., *ApJ*, **616** (2004) 357.
- [187] FROMANG S., BALBUS S. A., TERQUEM C. and DE VILLIERS J.-P., *ApJ*, **616** (2004) 364.
- [188] CAI K., DURISEN R. H., MICHAEL S., BOLEY A. C., MEJÍA A. C., PICKETT M. K. and D’ALESSIO P., *ApJ*, **636** (2006) L149.